Philipps-University
Faculty of Biology

Cloning of butyrogenic genes from *Clostridium difficile* and metabolic pathway reconstitution in *Escherichia coli*

vorgelegt von

El-Hussiny Ahmed Ahmed Ali Aboulnaga

aus Mansoura, Ägypten

Dissertation

zur

Erlangung des Doktorgrades

der Naturwissenschaften

(Dr. rer. nat.)

dem

Fachbereich Biologie

der Philipps-Universität, Marburg/Lahn

Deutschland

2011

Philipps-Universität
Fachbereich Biologie

Vom Fachbereich Biologie der Philipps-Universität Marburg als Dissertation am **26.07.2011** angenommen.

Erstgutachter

Prof. Dr.

Thorsten Selmer

Prof. für Biotechnologie &
Enzymtechnologie
Fachbereich Chemie & Biotechnologie,
Fachhochschule Aachen,
52428 Jülich,
Deutschland

Zweitgutachter

Prof. Dr.

Wolfgang Buckel

Prof. für Mikrobiologie &
Biochemie
Fachbereich Biologie,
Philipps-Universität,
D-35043 Marburg,
Deutschland

Tag der mündlichen Prüfung am **23.08.2011**

2011

Philipps-Universität
Fachbereich Biologie

Die Untersuchungen zur vorliegenden Arbeit wurden in der Zeit von November 2008 bis Mai 2011 am Fachbereich Chemie & Biotechnologie der Fachhochschule Aachen, Campus Jülich, unter der Leitung von **Prof. Dr. Thorsten Selmer** durchgeführt. Die Arbeit zudem wurde von **Prof. Dr. Wolfgang Buckel**, Phillips-Universität Marburg, und **Prof. Dr. Ahmed El-Refai**, Professor für Lebensmitteltechnologie der landwirtschaftlichen Fakultät der Mansoura Universität, Ägypten, wissenschaftlich betreut. Die vorliegende Arbeit ist mit Mitteln der ägyptischen Studien-Mission an **El-Hussiny Aboulnaga** unterstützt worden und wurde durch interne finanzielle Förderung der FH-Aachen (K2) gefördert.

2011

Philipps University
Faculty of Biology

Erklärung

Hiermit versichere ich, dass ich meine Dissertation mit dem Titel

"Cloning of butyrogenic genes from *Clostridium difficile* and metabolic pathway reconstitution in *Escherichia coli*"

selbständig, ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen und Hilfen bedient habe. Die vorliegende Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen deutschen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg, Aboulnaga, El-Hussiny

(Ort, Datum) (Unterschrift)

2011

ACKNOWLEDGMENT

Firstly, great and foremost thanks to "**ALLAH**" for making completion of this work.

I am very grateful to **Prof. Dr. Wolfgang Buckel**, who offered me the opportunity to carry out this thesis under his supervision in Philipps-University Marburg, for his valuable supervision, and generous support throughout this work.

I would like to show my gratitude to **Prof. Dr. Thorsten Selmer**, who accepted me in his laboratory to complete this work, for generous planning, for expert supervision, remarkable correction of the thesis, and his undeniable help throughout this work.

I would like to express my appreciation to **Prof. Dr. Ahmed El-Refai** Professor of Food Science, Mansoura University for his continuous support and encouragement, offering valuable suggestions.

I am grateful also to my colleague **Johannes Schiffels** for helping me to measure short-chain fatty acids with HPLC as well helpful discussions.

I also appreciate the excellent technical assistance from **Brigitte Lehan** and **Melanie Classen**.

I wish to express my thanks to **Regina Fischer**, who kindly provided some vectors and expression clones used in this study.

Heartily thanks are due to all the **staff members** of the Department of Food industries, Faculty of Agriculture, Mansoura University, Egypt, whose gave me chance to complete my PhD degree in Germany, special regards for all of them.

I wish to express my deep thanks to my family particully my **parents** and my small kids, **Youssuf** and **Nour El-Din**.

Finally, this work would not have been possible without the co-operation and support of my lovely wife **Al-Zahraa**, special thanks to you.

Curriculum vitae

Personal data

Name:	**El-Hussiny Ahmed Ahmed Ali Aboulnaga**
Address:	**Heinrich-Mußmann Str. 1, 52428 Campus Jülich, Germany**
Date of birth:	**28.06.1980**
City:	**Mansoura, Egypt**
Nationality:	**Egyptian**
E-Mail:	**h_aboulnaga@yahoo.com** **h_aboulnaga@anut.fh-aachen.de**

Education

1995-1997	Higher Secondary School.
1998-2001	Bachelor of Science (Faculty of Agriculture, Mansoura University, Egypt)
2002-2005	Master of Science (Food Industries Department, Faculty of Agriculture, Mansoura University, Egypt). Subjects "Chemical and technological studies on some occurring naturally anti-enzymes in foods".
2005-2008	An assistant lecturer in Food Industries Dept., Faculty of Agriculture, Mansoura University, Egypt.
Nov. 2008-till now	PhD student in the research group of Prof. Dr. **Thorsten Selmer**, Biotechnology institute, FH-Aachen, 52428 Campus Jülich, Germany.

Publications

Orginal paper:

- **Aboulnaga, E., O. Pinkenburg, J. Schiffels, A. El-Refai, W. Buckel and T. Selmer. (2011).** *Clostridium difficile* takes advantage of butyryl-CoA dehydrogenase driven electron transport phosphorylation. (*In preparation)*

Conferences:

- **Aboulnaga, E., A. El-Refai, W. Buckel and T. Selmer. (2009).** Cloning and expression of *Clostridium difficile* butyrate synthetic genes in *Escherichia coli*. **2nd Graduate Symposium**: November 5th, 2009. FH Aachen –University of Applied Sciences, Campus Jülich, Germany. (Abstract and poster)

- **Aboulnaga, E., J. Schiffels, W. Buckel and T. Selmer. (2010).** Metabolic pathway design: Transfer of a butyrate forming module from *Clostridium difficile* to *Escherichia coli*. **3rd Graduate Symposium**: November 5th, 2009. FH Aachen – University of Applied Sciences, Campus Jülich, Germany. (Abstract and poster)

Abbreviations

AAD	Antibiotic-Associated Diarrhea
AdhE	Bi-functional aldehyde/ alcohol dehydrogenase
AHT	Anhydrotetracycline
Amp^r	Ampicillin resistance
APS	Ammonium Persulfate
ATP	Adenosine-5'-triphosphate
BCD	Butyryl-CoA dehydrogenase
BHBD	β-Hydroxybutyryl-CoA dehydrogenase
BUK	Butyrate Kinase
CAI	Codon Adaptation Index
CDM	Cell Dry Mass
CDSs	Coding Sequences
Cm^r	Chloramphenicol resistance
CRT2	Crotonase
dNPs	Deoxy-Nucleotide Tri-phosphates
DTB	Desthiobiotin
DTT	1,4-Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
Etf	Electron transfer flavoprotein
ETP	Electron Transport coupled to Phosphorylation
FAD	Flavin Adenine Dinucleotide
$Fc^+FP_6^-$	Ferricenium hexafluorophosphate
Fd_{ox}	Ferredoxin oxidized form
Fd_{red}	Ferredoxin reduced form
FPR	Ferredoxin:NADP reductase
G-6-P DH	Glucose-6-phosphate Dehydrogenase
GOI	Gene Of Interest

GUK	Glucokinase
HABA	Hydroxy-azophenylbenzoic acid
HCDC	High Cell Density Cultivation
Kan^r	Kanamycin resistance
NAD^+	Nicotinamide Adenine Dinucleotide
$NADP^+$	Nicotinamide Adenine Dinucleotide Phosphate
OD	Optical Density
PBT	Phosphate butyryl-transferase
PCR	Polymerase Chain Reaction
q_B	Specific butyrate production rate
q_G	Specific glucose consumption rate
q_G^{max}	Maximum Specific glucose consumption rate
RBS	Ribosomal Binding Site
RFLP	Restricts-Fragment-Length-Polymorphism
Rnf	Rhodobactor Nitrogen Fixation
SDS	Sodium Dodecyl Sulfate
SLP	Substrate Level Phosphorylation
TEMED	N, N, N', N'- Tetraethylethylenediamine
THIA1	Thiolase
X-Gal	5-Bromo-4-chloro-3-indolyl β-D-galactoside
$Y_{B/G}$	Yield of butyrate per mmol glucose consumed
$Y_{B/X}$	Yield of butyrate per mmol- biomass
μ	Specific growth rate
μ^{max}	Maximum specific growth rate
ε	Extinction coefficient

List of Tables

List of Figures

Summary

In this study, a novel strategy for systematic applications in synthetic biology aiming the transfer of metabolic capacities from a donor to an acceptor organism was elucidated. As a model system, it was chosen the butyrate fermentation of *Clostridium difficile*, which has not been studied in details before. In order to succeed, individual enzymes, thiolase (EC 2.3.1.9), β-hydroxybutyryl-CoA dehydrogenase (EC 1.1.1.157), crotonase (EC 4.2.1.17), butyryl-CoA dehydrogenase with its two electron transferring flavoprotein subunits (EC 1.3.99.2), phosphate butyryltransferase (EC 2.3.1.19) and butyrate kinase (EC 2.7.2.7), were initially produced in *E. coli* in order to demonstrate their functional production, followed by an initial characterization *in vitro*. Substrate specificity, stability and kinetic parameters were established in order to address their capability for the synthesis of various metabolic intermediates in volatile fatty acid biosynthetic pathways. In this context, novel enzyme tests for phosphate butyryl-transferase and butyrate kinase were developed and a facile method for testing of bi-forking butyryl-CoA dehydrogenase in the physiological reaction was established. Using the latter assay, evidence was obtained for a bi-forking butyryl-CoA dehydrogenase in *C. difficile*, which could allow energy conservation aided by reduced ferredoxin, an ability lacking in the most intensively studied pathway of *C. acetobutylicum*. The impact of such enzymes, and of the second rate limiting one (3-hydroxybutyryl-CoA dehydrogenase), in butan-1-ol-forming recombinant strains is discussed.

Successively, the enzyme encoding genes were assembled in a synthetic operon in order to create an artificial metabolic pathway module, which was expressed from a single plasmid and enabled the formation of the predicted product, butyrate. Two different butyrate metabolic pathway modules, *pASG.wt_BUT*$_D$ and *pASG.wt_BUT*$_{AC}$, were constructed, which contained the butyrate fermentation genes (*thia*$_1$, *hbd*, *crt*$_2$, *pbt* and *buk*) from *C. difficile*, but differed in the origin of the butyryl-CoA dehydrogenase/ETF complex genes. These genes were either from *C. difficile* (*pASG.wt_BUT*$_D$) or *C. acetobutylicum* (*pASG.wt_BUT*$_{AC}$).

The recombinant butyrogenic pathways were analyzed *in vivo* by a combination of enzymes activity measurements in cell free extracts and analysis of the fermentation process: substrate uptake, biomass-, product- and by-product- formations were established in order to understand the mass flows in the process. Only the recombinant strains were capable to produce butyrate with a maximum concentration of 3.1 and 3.7 mM, respectively. While growth characteristics, substrate consumption and product yields were rather similar in the recombinant strains, the butyryl-CoA dehydrogenase of different origin and sub-class caused remarkable different kinetics of butyrate formation, which are discussed with regard to their potential applications in solventogenis.

Zusammenfassung

Die vorliegende Arbeit stellt die Entwicklung eines neuartigen Modellsystems dar, welches die Implementierung kompletter Stoffwechselmodule in geeignete Akzeptororganismen zur systematischen Anwendung in der synthetischen Biologie erlaubt. Das bisher wenig untersuchte butyrogene Stoffwechselmodul aus *Clostridium difficile* wurde hierzu in *Escherichia coli* transferiert, bestehend aus den individuellen Enzymen Thiolase (EC 2.3.1.9), β-Hydroxybutyryl-CoA-Dehydrogenase (EC 1.1.1.157), Crotonase (EC 4.2.1.17), Butyryl-CoA-Dehydrogenase einschließlich der zwei Elektronen transferierenden Flavoprotein-Untereinheiten (EC 1.3.99.2), Phosphotransbutyrylase (EC 2.3.1.19) sowie Butyratkinase (EC 2.7.2.7). Die Enzyme wurden zunächst einzeln in *E. coli* produziert, funktionell getestet und *in vitro* charakterisiert. Die dabei erhaltenen kinetischen Parameter sowie Daten zur Substratspezifität und Stabilität gaben Aufschluss über die Fähigkeit der Enzyme zur Produktion von Intermediaten innerhalb von Modulen zur Produktion kurzkettiger Fettsäuren. In diesem Kontext wurden neue Aktivitätstests für Phosphotransbutyrylase und Butyratkinase etabliert. Im Zuge der Entwicklung einer vereinfachten Methode zur Messung der verzweigten Butyryl-CoA-Dehydrogenase in der physiologischen Reaktionsrichtung konnte die Annahme erhärtet werden, das *C. difficile* entgegen dem intensiv untersuchten *C. acetobutylicum* mithilfe dieses Enzyms zur Energiekonservierung mit reduziertem Ferredoxin fähig ist. Der potentielle Einfluss derartiger Enzyme bei der Produktion von Butan-1-ol in rekombinanten Stämmen wird in diesem Kontext diskutiert.

Zur Konstruktion des kompletten artifiziellen Stoffwechselmoduls wurden die Gene für oben erwähnte Enzyme erfolgreich zu einem synthetischen Operon assembliert, resultierend in den Expressionsplasmiden *pASG.wt_BUT*$_D$ und *pASG.wt_BUT*$_{AC}$, welche sich lediglich in der Herkunft des Butyryl-CoA-Dehydrogenase/ETF-Komplexes unterscheiden. Die Gene für den besagten Komplex stammten entweder aus *C. difficile* (*pASG.wt_BUT*$_D$) oder *C. acetobutylicum* (*pASG.wt_BUT*$_{AC}$).

Die rekombinanten Butyrogenen Stoffwechselmodule wurden *in vivo* untersucht durch Kombinationen aus Enzymtests in zellfreien Extrakten und die Analyse des Fermentationsprozesses im Bezug auf Substrataufnahme, Biomasse-, Produkt- sowie Nebenproduktbildung zum Verständnis des Massenflusses innerhalb des Systems. Tatsächlich waren lediglich die rekombinanten Stämme zur Butyrat-Produktion fähig, mit maximalen Konzentrationen von 3,1 bzw. 3,7 mM im Medium. Während keine großen Unterschiede zwischen den beiden rekombinanten Stämmen bezüglich Wachstum, Substratverbrauch und Produktausbeute beobachtet werden konnten, hatte die Herkunft der jeweiligen Butyryl-CoA-Dehydrogenase dramatische Auswirkungen auf die Kinetik der Produktbildung, eine Beobachtung, die besonders im Bezug auf den potentiellen Einsatz in der biotechnologischen Lösemittelproduktion diskutiert wird.

1. Introduction

1.1. Clostridia

1.1.1. *Clostridium difficile*

Clostridia are strict anaerobic, spore-forming rods (bacilli), typically rod-shaped, occurring mainly in soil, but also in the intestines of humans and animals **[1]**. The genus *Clostridium* includes human pathogenic species causing botulism, tetanus gas gangrene and the like **[1]**. Clostridial endospores are heat resistant and can survive in the environment for years despite desiccation and exposure to disinfectants. There are many reports on viable spores being found on clothing, hospital furniture and equipment. Thus, these spores contribute to exogenous exposure of humans via the fecal-oral route (food-borne diseases like botulism or *C. difficile*-associated diarrhea) or wound infections (tetanus). Spores and toxins can survive ingestion and vegetative cells capable of growth and toxin production can proliferate in the intestine or wound tissues **[2]**.

Clostridium difficile is a species of Gram-positive bacteria of the genus *Clostridium* that causes diarrhea, colitis and other intestinal disease when competing bacteria are wiped out by antibiotics (Antibiotic-Associated Diarrhea, AAD) **[2-5]**. Approximately 15% to 20% of cases of AAD are caused by *C. difficile* **[6, 7]**, which is, therefore, an emerging human pathogen in hospitalized environments. Originally isolated in 1935, *C. difficile* was initially thought to be a component of normal gut flora, and was not identified as a pathogen until the 1970s, when colitis associated with clindamycin treatment was further investigated **[2]**. Pathogenic strains of *C. difficile* cause diarrhea and colitis via toxin production **[8]**. Two major toxins have been identified: toxin A is a 308 kDa enterotoxin, and toxin B is a 269 kDa cytotoxin. Both toxins are capable of stimulating production of proinflammatory cytokines that have been implicated in the pathogenesis of pseudomembranous colitis **[7, 8]**.

In the mid-2010th, there are 14 clostridia genomes at different sequencing stages, but only three, *C. acetobutylicum* ATCC 824 (cluster I, subcluster Ib), the non-motile *C. perfringens* strain 13 (cluster I, subcluster Ib) and the sporogeneous *C. tetani* E88 (cluster I, subcluster Ie), have been annotated and made public by NCBI **[9]**. The genomes of *C. difficile* strain 630 (cluster XI) and *C. botulinum* Hall strain A (ATCC 3502) (cluster I, subcluster Id) have been sequenced and all gaps closed **[9, 10]**. The putative open reading frames sequences for *C. difficile* can be downloaded from the Sanger center **[9-11]**. The genome of *C. difficile* strain 630 consists of a circular chromosome of 4,290,252 bp and a plasmid, pCD630, of 7,881 bp. The chromosome encodes 3,776 predicted coding sequences

(CDSs); in common with other low GC content Gram-positive bacteria, the chromosome has a strong coding bias, with 82.1% of the CDSs encoded on the leading strand. The plasmid carries 11 CDSs, none of which has any obvious function. Reciprocal FASTA analysis against four annotated clostridial genomes, *C. acetobutylicum* **[12]**, *C. botulinum* **[13]**, *C. perfringens* **[14]**, and *C. tetani* **[15]**, showed that about 913 (25%) of *C. difficile* CDSs are shared with all the sequenced clostridia, whereas 2,065 (56%) are unique to *C. difficile* (Fig.1.1) **[9, 10]**. The conserved clostridial CDSs mainly encode essential functions, whereas the *C. difficile* unique CDSs encode many accessory functions and mobile elements **[10]**. The availability of genomic information for a range of *C. difficile* strains affords researchers the opportunity to better understand not only the evolution of these organisms but also their basic physiology and biochemistry **[16]**.

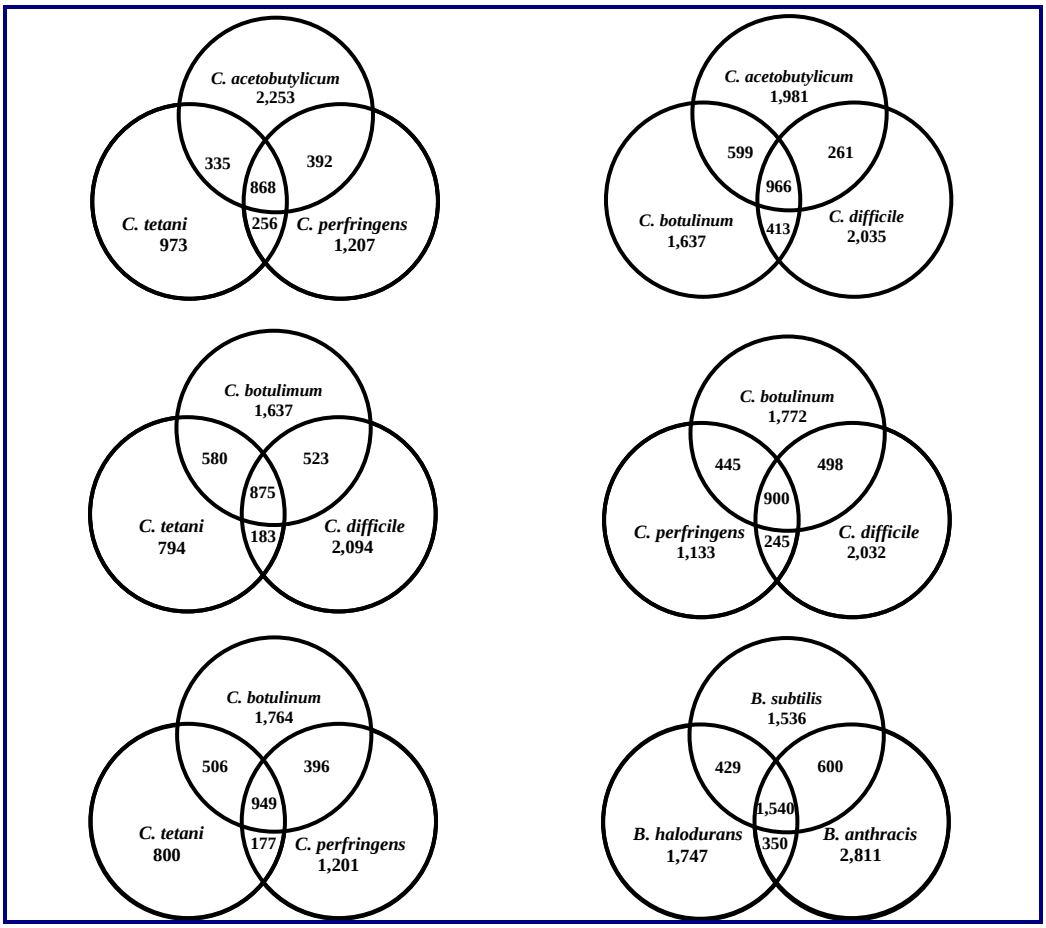


Fig.1.1: The number of putative orthologous genes in different clostridia; *C. acetobutylicum*, *C. tetani*, *C. perfringens* strain 13, *C. botulinum*, and *C. difficile*. For comparison, the number of putative orthologous genes between *Bacillus subtilis*, *B. halodurans* and the *B. anthracis* Ames strain is shown. In every Venn diagram, the genome on top was used as a reference for determining the number of putative orthologs. For comparing the bottom genomes of each Venn diagram, the genome on the right-hand side was used as a reference **[9, 15]**.

Data in figure (1.1) show that generally three different given clostridia share between 850 and 950 proteins. This number is lower than the ~1,550 proteins shared between *Bacilli*, (perhaps) reflecting the heterogeneity of the *Clostridium* genus **[9]**. The first comparison of putative orthologous pairs among *C. tetani*, *C. perfringens*, and *C. acetobutylicum* was done by ***Brueggemann*** and co-workers and they observed that orthologous genes were defined as best–best matches having >30% amino acid identity over >60% of the length of both genes **[15]**. ***Paredes*** and co-workers noted that the solventogenic species *C. acetobutylicum* and the pathogen *C. botulinum* share the highest number of putative orthologous pairs (1,565), even though they belong to different phylogenetic subclusters. In general, *C. botulinum* shares the highest number of putative orthologs with any other analyzed clostridial species, regardless of their pathogenicity or phylogenetic classification **[9]**.

1.1.2. Butyrate pathway in clostridia

Anaerobic bacteria produce crotonyl-CoA, which is the precursor of butyrate, either by amino acid fermentation e.g. from glutamate, lysine, threonine and methionine or by synthesis from two acetyl-CoA molecules derived from glucose or glutamate and other amino acids via e.g. 3-methylasparate pathway **[17]**.

The metabolic pathway of glucose in acidogenic clostridia has several possible end-products including butyrate and acetate in addition to CO_2, H_2 and lactate as by-products. There are six enzymes involved the conversion of acetyl-CoA to butyrate (Fig.1.2): Thiolase (EC 2.3.1.9) **[18-23]**, β-hydroxybutyryl-CoA dehydrogenase (EC 1.1.1.157) **[24-27]**, crotonase (EC 4.2.1.17) **[25, 28-32]**, butyryl-CoA dehydrogenase/ETF complex (EC 1.3.99.2) **[25, 33-37]**, phosphate butyryltransferase (EC 2.3.1.19) **[38-42]** and butyrate kinase (EC 2.7.2.7) **[43-47]**. The butyrate biosynthetic pathway starts with condensation of two molecules of acetyl-CoA by thiolase to acetoacetyl-CoA which is reduced to β –hydroxybutyryl-CoA involving a dedicated dehydrogenase. Then, water is eliminated via crotonase and the resulting crotonyl-CoA is reduced to butyryl-CoA by the butyryl-CoA dehydrogenase complex composed of three individual polypeptides. Butyryl-CoA is further converted by phosphate butyryltransferase and the butyrylphosphate thus formed is subsequently used for substrate phosphorylation of ADP, mediated by butyrate kinase.

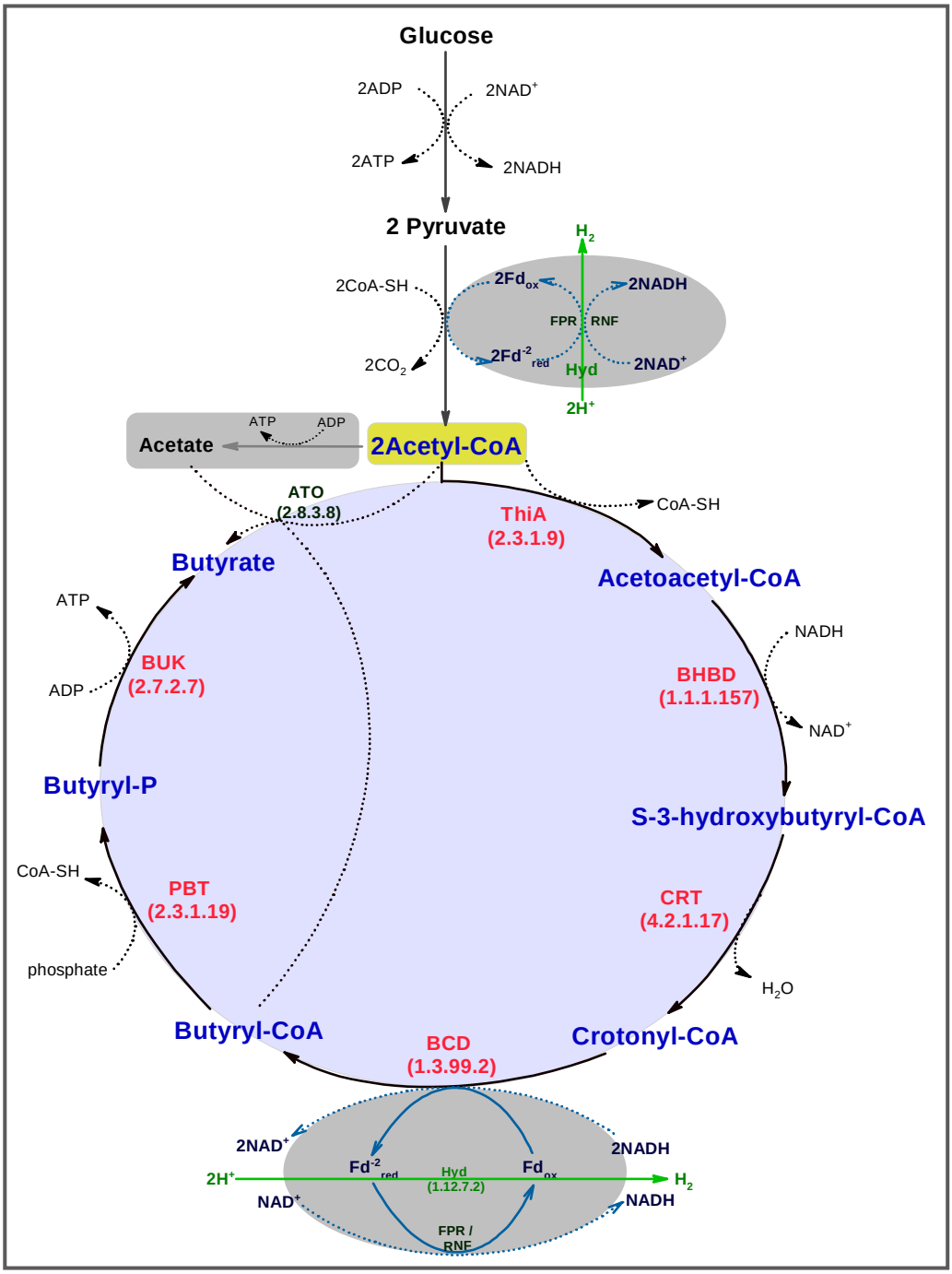


Fig.1.2: The butyrate biosynthetic pathway in clostridia fermentations of glucose.

1.1.3. Energy conservation by butyryl-CoA dehydrogenase

In bacteria, there are two basic mechanisms by which redox reactions are coupled to energy conservation: substrate level phosphorylation (SLP) and electron transport coupled to phosphorylation (ETP). The majority of organisms use oxygen as the terminal acceptor, but anaerobic bacteria are able to use other electron acceptors such as nitrate, nitrite, sulfate, carbon dioxide, fumarate, the substrate itself or a derivative thereof to respire **[17]**. Many redox processes cause proton extrusion from the cytoplasm and therefore contribute to an

electrochemical potential which enables energy conservation in a respiratory-like process for that fermentation cannot be completely separated from a membrane electron transport **[48]**.

Recently, NADH-dependent reduction of crotonyl-CoA to butyryl-CoA in butyrate synthesis via electron bifurcation butyryl-CoA dehydrogenase/ETF complex and ferredoxin has been shown (Fig.1.3) **[17, 33]**. The reduction of crotonyl-CoA (E^0= -10 mV) with NADH (E^0= -320 mV) is highly exergonic and occurs spontaneously. However, the reduction of ferredoxin (E^0= -410 mV) with NADH is endergonic and the process cannot occur **[17, 33]**. Therefore, the only possibility to reduce ferredoxin is to drive its endergonic reduction by the exergonic reduction of crotonyl-CoA, which makes the overall reaction exergonic (ΔG^0= -40 kJ.mol^{-1}) **[17]**. In the coupled reactions, BCD/ETF complex bifurcates the two electrons from NADH; one electron proceeds to the more positive electron acceptor butyryl-CoA dehydrogenase and finally to crotonyl-CoA, and the other is transported to the more negative acceptor ferredoxin (Fd ⟶ Fd^-). The next NADH delivers the second electron to complete the reduction of crotonyl-CoA to butyryl-CoA and Fd^- **[17, 33]**.

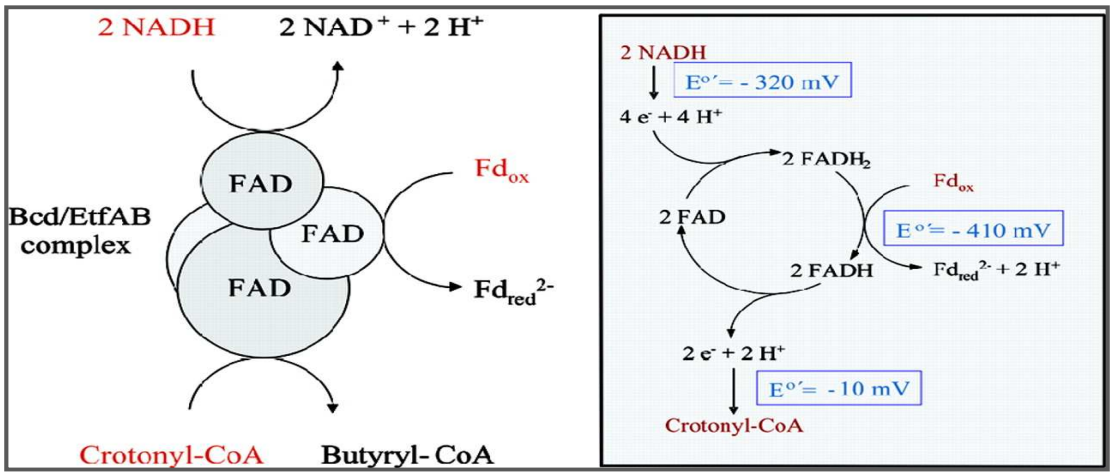


Fig.1.3: ***Li*** **and co-worker proposed the mechanism of endergonic ferredoxin reduction with NADH coupled to exergonic crotonyl-CoA reduction with NADH catalyzed by the butyryl-CoA dehydrogenase/ETF complex from** ***C. kluyveri*** **[33].**

The reduced ferredoxin generated by BCD/ETF gets reoxidized by the membrane-bound ferredoxin:NAD oxidoreductase complex (Rnf-CDGEAB), which utilizes the exergonic ferredoxin oxidation to establish a proton motive force across the cell membrane **[17, 33]**. Thereby, additional energy conservation via V-type ATPases is possible **[49]**. Such a system has been proposed to improve the ATP yield in the fermentation of glutamate via 3-methylaspartate up to 0.9 mol ATP per mol glutamate in *Clostridium tetanomorphum*: 5 mol glutamate yield 3 mol ATP via SLP and 1.5 ATP equivalents (6 * 0.25) by the formation of 6 NADH via Rnf **[17]**.

Furthermore, there are two other mechanisms to reoxidize reduced ferredoxin, either by NADP:ferredoxin oxidoreductase (FPR, EC 1.18.1.2), which couples the oxidation of reduced ferredoxin with $NADP^+$ reduction, or ferredoxin-hydrogenase (Hyd, EC 1.12.7.2) which couples hydrogen production with ferredoxin oxidation. Based on the genome-derived data from different clostridia, there is at least one system to reoxidize reduced ferredoxin present in species (Figure 1.4). Some organisms such as *C. difficile* and *C. kluyveri* contain all of the three systems.

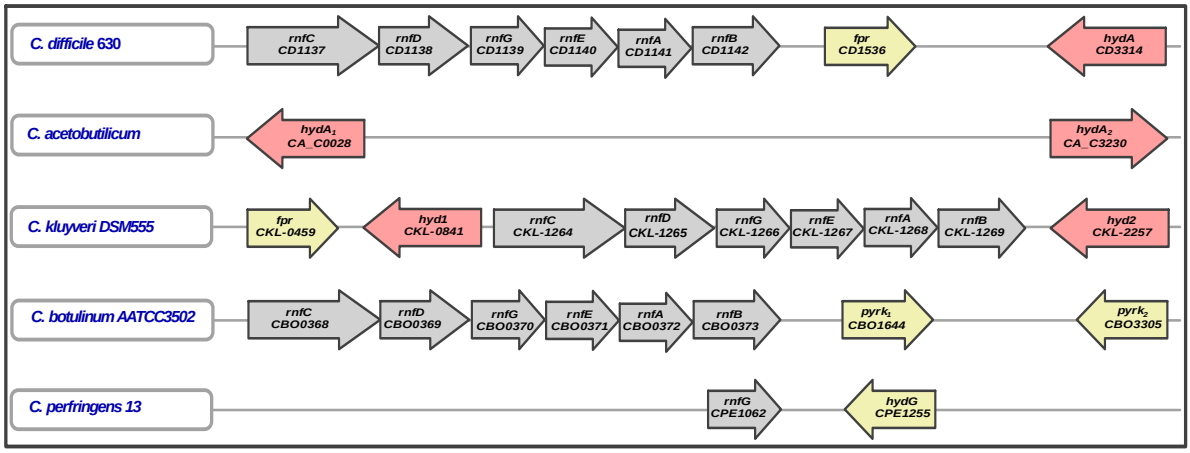


Fig.1.4: Genome analysis for genes encoding proteins capable to reoxidize reduced ferredoxin in different clostridia. There is in each arrow the name and the accession number for individual genes. The direction of arrow indicates the open reading frame direction. Linked arrows show that these genes are clustered in one operon. The arrow length correspond to the genes sizes.

1.1.4. Comparative analysis of butyrate biosynthetic genes in clostridia

The arrangement of homologous butyrate biosynthetic pathway genes in the genome of *C. difficile* 630, *C. acetobutylicum*, *C. kluyveri* DSM555, *C. botulinum* AATCC3502 and *C. perfringens* 13 are shown in figure (1.5). The data shows that there are 5 genes (*hbd*, *crt*, *bcd*, *etfB* and *etfA*) encoding for 3-hydroxybutryl-CoA dehydrogenase, crotonase, butyryl–CoA dehydrogenase and the two electron transfer flavoprotein subunits, respectively. They are clustered together in one operon in those species. Moreover, some species like *C. difficile* and *C. botulinum* have the gene encoding for thiolase (*thia*) clustered in the above operon and this indicates that in these two species the genes encoding enzymes capable of converting acetyl-CoA to butyryl-CoA are transcribed under a single regulator. Also the analysis shows that the genes encoding for phosphate butyryl-transferase (*pbt*) and butyrate kinase (*buk*) are transcribed in another operon. This of course signalizes the ability to switch between acid and alcohol production, since the butyryl-CoA is the precursor in both acid and

alcohol so the organism should separate the control of butyryl-CoA production and its subsequent conversation to acid or alcohol. In the case of *C. kluyveri* the lack of *pbt-buk* operon is compensated by another gene (*cat*$_3$, accession number CKL_3595) encoding for abroad substrate specificity: acetate-CoA transferase that affords the formation of butyrate and caproate **[50]**. It is worth mentioning that some of these enzymes are found in the clostridia genomes in paralogs, which do, however, not cluster and were, therefore, omitted from the overview figure1.5.

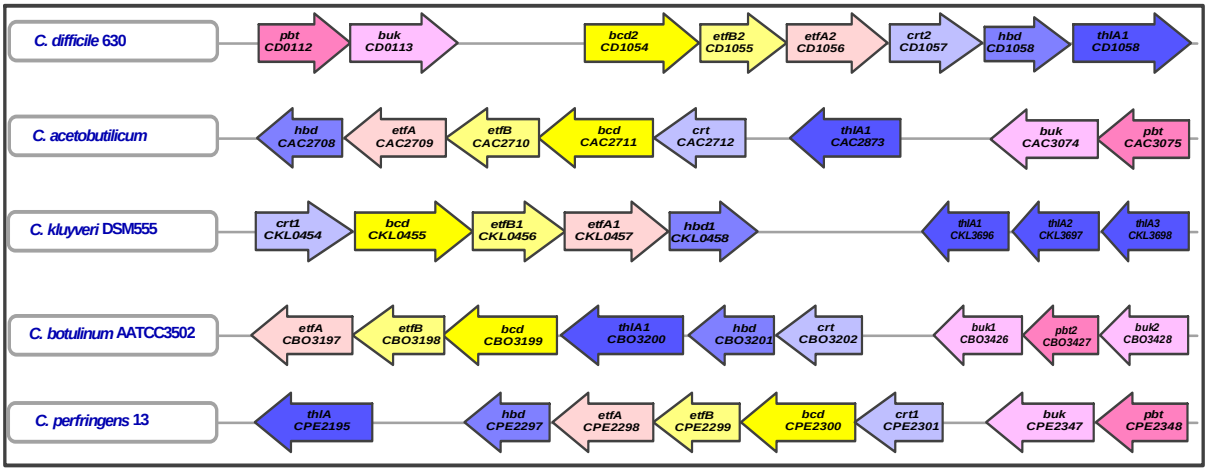


Fig.1.5: The arrangement of homologous butyrate biosynthetic pathway genes in different clostridia. There is in each arrow the gene name plus the gene accession number. The direction of arrow indicates the open reading frame direction. Linked arrows show that these genes are transcript in one operon; however the unlinked genes are not clustered. The arrow sizes are corresponding to the genes sizes.

1.2. *Escherichia coli*

Escherichia coli is a species of Gram negative, rod-shaped, facultative anaerobic, non-sporulating bacterium that is commonly found in the lower intestine of warm-blooded organisms. *E. coli* was discovered by German pediatrician and bacteriologist ***Theodor Escherich*** in 1885 **[51]**, and is now classified as part of the Enterobacteriaceae family of γ-proteobacteria **[52]**. Most *E. coli* strains are harmless, part of the normal flora of the gut and can benefit their host by producing vitamin K_2. However, some *E. coli* strains such as serotype O157:H7 can cause serious food poisoning in humans **[53]**.

1.2.1. Advantages of using *E. coli*

E. coli has been the most widely used prokaryotic host for mass-production of recombinant proteins of pharmaceutical and industrial purposes. This remarkable microorganism has numerous desirable characteristics as a production host such as fast cell

growth, easy genetic manipulation, straightforward high cell density cultivation (HCDC) and capacity to hold over 50% of foreign protein in total protein expression **[54, 55]**. These features have facilitated the production of recombinant proteins and non-protein bio-molecular products such as amino acids, primary and secondary metabolites with high productivities **[54, 56-64]**. In addition, HCDC has the added advantage of increased cost effectiveness, reduced culture volume, enhanced downstream processing, reduced wastewater, lower production costs and a reduced investment in equipment. The effectiveness of HCDC can be further improved by employing a host strain optimized by various strategies. The most popular approaches used are the use of different promoters to regulate expression levels, different host strains, co-expression of various proteins (e.g. chaperones), reduction of culture temperature and secretion of proteins into the periplasm or culture medium **[55, 65]**.

1.2.2. Different metabolic pathways transferred into *E. coli*

Because clostridia are more difficult to transform, sensitive to oxygen, exhibit complex fermentation kinetics and can form spores, they are more difficult to engineer. Synthetic biology promises to simplify and expedite the design and engineering of metabolic pathways for the bioconversion of various feedstocks to desired fuels and chemicals. For that, several groups have recently applied synthetic biology approaches to the synthesis of desired fuels and chemicals in a well-studied, robust host organism such as *E. coli* **[56-59, 66, 67]**, see table (1.1).

Table.1.1: Some metabolic pathways transferred into *E. coli* [68]

Source	Transferred genes	Substrate	Major-Product	Titer	Reference
C. acetobutylicum	*thiA, hbd, crt, bcd-etfB-etfA, and adhe1*	Glucose	Butan-1-ol	1.2g/L 0.5g/L	**[57]** **[59]**
Methanococcus jannaschii	CimA, leuABCD	Glucose	1-Propanol & 1-Butanol	3.5g/L 0.5g/L	**[69]**
Different organisms	Combinatorial of different genes	Glucose	Isopropanol	4.9g/L	**[64]**
Zymomonas mobilis	*pdc, adhA, and adhB*	Xylose	Ethanol	40g/L	**[70]**
Bacillus subtilis	*yhfR and nudF*	Glucose	Isopentenol	0.11g/L	**[71]**
Different organisms	*dhaB1, pduP, phaC1*	Glycerol	poly(3-hydroxypropionate)	1.42g/L	**[72]**
C. acetobutylicum *Ralstonia eutropha*	*hbd, thl, ptb and buk* + *bktB, phaA and phaB*	Glucose	(R)- 3-Hydroxybutyrate & (S)-(3HB)	2.92g/L & 2.08g/L	**[60]**

1.2.3. Metabolic pathway design in *E. coli*

The analysis of metabolic activities in bacteria is a useful tool for the generation and development of novel strains with desired growth and production properties. Several methods are available for the evaluation and understanding of the metabolic pathways: Biochemical methods to measure metabolite concentrations and enzyme activity **[56, 57, 59, 66, 69, 73]**, mathematical models to analyze metabolic fluxes through the various pathways **[74-78]**, proteomic studies to identify expressed proteins **[79-83]**, and molecular biology methods to detect and analyse gene expression **[80, 84-86]**. A combination of various methods is required to obtain a comprehensive understanding of metabolic activities. In particular high-throughput methods provide substantial amount information on global gene expression and, thereby offer routes to system biological approaches, but do not always reflect the actual activity of the individual components.

Biochemical methods can provide additional data measuring either metabolite concentrations or enzyme activities. These measurements were used to analyze the metabolic differences between *E. coli* K (strain JM109) and *E. coli* B (BL21) to solve the growing behavior puzzle between these strains. It was established that when growing at high glucose concentration (20-40 g/L), *E. coli* B (BL21) grew faster, to higher densities and produced less acetate than *E. coli* K (JM109). The metabolite analyses showed that *E. coli* B (BL21) had a higher internal concentration of isocitrate and a lower concentration of pyruvate. In regard to the enzymatic activity, isocitrate dehydrogenase and isocitrate lyase highly active in *E. coli* B (BL21) compared with almost no detectable activity of isocitrate lyase in *E. coli* K (JM109) **[87]**.

Regarding the transferred pathway activities in *E. coli*, ***Fischer*** and co-worker developed another application for the determination of the intracellular pathway intermediates to assess the heterologous butyrate and butan-1-ol pathway activities in recombinant *E. coli* strains. They observed that there are a two rate-limiting steps mediated by two individual enzymes, 3-hydroxybutyryl-CoA dehydrogenase and butyryl-CoA dehydrogenase/ ETF complex, and increasing the expression level of these enzymes alleviated the bottleneck in this pathway **[66]**. Other groups were satisfied by measuring the enzyme activities in cell free extracts obtained from recombinant *E. coli* strains harboring butanol-forming genes from clostridia in addition to measure butan-1-ol concentration in the culture. They also observed that butyryl-CoA dehydrogenase/ ETF complex is a rate-limiting enzyme in this pathway **[57, 59]**.

Regarding mathematical methods, ***Lequeux*** group applied a dynamic metabolic flux analysis on *E. coli* cultivations shifting from carbon limitation to nitrogen limitation and vice versa and they observed a lag phase after switching the limiting substrate from nitrogen to carbon accompanied with an increase in maintenance energy requirement. However, this lag phase did not occur upon C- to N-limitation switch **[88].**

1.2.4. Fatty acid oxidation in *E. coli*

Fatty acids and their derivatives play essential roles in a number of cellular processes, including cell signaling, transcriptional control, membrane syntheses and protein modification. Thus, fatty acid metabolism is tightly regulated to allow the cells to adapt quickly to environmental changes. *E. coli* can uptake and utilize long-chain fatty acids (C_{12} to C_{18}) as a sole carbon and energy source **[89-91]**. In contrast, it cannot uptake but utilize the medium-chain fatty acids (C_6 to C_{11}), and neither uptake nor utilize short-chain fatty acids (C_4 and C_5) **[89, 90]**.

The uptake and utilization of fatty acids occurs by means of proteins encoded by the *fad* regulon genes. The *fad* regulon encoded proteins are essential for fatty acids transport, activation and β-oxidation. Exogenous fatty acids are transported into the cell by the outer-membrane-bound fatty acid transport protein FadL and activated by the inner-membrane-associated acyl-CoA synthetase FadD yielding acyl-CoA thioesters. Then, the activated fatty acids are catabolized to acetyl-CoA via β-oxidation mediated by the acyl-CoA dehydrogenase encoded by *fadE* and the tetrameric enzyme complex encoded by *fadB* and *fadA*. Acetyl-CoA is subsequently used to generate energy and precursors required for cell maintenance **[92]**.

1.3. Aim of investigation

In the recent years, several metabolic pathways were transferred into *E. coli* for the bioconversion of various feedstocks to desired fuels and chemicals using common metabolic pathway engineering pathways. These ways started with transferring the metabolic pathway genes into *E. coli* using more than one expression plasmid with different antibiotic resistances and sometimes different regulators. Then, the transferred metabolic pathway in *E. coli* was analyzed using either biochemical, mathematical, proteomic or genomic methods. Our goal was to design a new strategy for the evaluation and understanding of the transferred metabolic pathways in *E. coli*. The novel strategy must allow i) rapid recombinant production of dedicated enzymes in a functional state for individual testing, ii) systematic assembly of individually characterized enzyme-encoding genes to yield functional metabolic pathway

modules and iii) the tailor-made design of technical microbes for different applications from these metabolic modules.

The target genes for introduction in *E. coli* were chosen from the genome sequenced human pathogen *Clostridium difficile*. The butyrate formation by this organism has not yet been studied on the enzyme level and recent findings suggest that the butyrate pathway might be crucial for general understanding of clostridial metabolism and might be of particular importance in understanding of the remarkable metabolic versatility of *C. difficile*.

In order to test the capability of a recently developed novel combinatorial cloning system for these applications, we decided to transfer the capability to form butyrate from *Clostridium difficile* to *Escherichia coli*, which naturally lacks the ability to convert glucose to butyric acid. In a systematic approach, thiolase (EC 2.3.1.9), β-hydroxybutyryl-CoA dehydrogenase (EC 1.1.1.157), crotonase (EC 4.2.1.17), butyryl-CoA dehydrogenase/ ETF complex (EC 1.3.99.2), phosphate butyryltransferase (EC 2.3.1.19) and butyrate kinase (EC 2.7.2.7) were first produced in *E. coli* and functionally tested *in vitro*. Substrate specificity, stability and kinetic parameters were established in order to address their capability for the synthesis of various metabolic intermediates in volatile fatty acid biosynthetic pathways. Then, the individual genes were combined in order to yield a synthetic metabolic pathway module, which was introduced into the host cells. The transferred pathway was analyzed by measuring the enzyme activities in cell free extracts and analysis of the fermentation process: substrate uptake, biomass-, product- and by-product formations were established in order to understand the mass flows in the process.

2. Materials

2.1. Chemicals, biochemicals and reagents

Unless otherwise stated, all chemicals were purchased from Sigma-Aldrich (Taufkirchen, Germany), VWR International GmbH (Darmstadt, Germany) or Roth (Karlsruhe, Germany) and were of highest available quality.

2.2. Chemicals synthesis

2.2.1. Chemical synthesis of Butyryl-phosphate

Butyryl-phosphate was prepared according to ***Uede*** **[93]** with slight adaptations: To a mixture of 2.39 mL of phosphoric acid (85%) and 18.37 mL butyric acid, 34.19 mL butyric anhydride were added dropwise and allowed to stir overnight at room temperature. Then, 1.7 g of solid LiOH was added for product precipitation at room temperature for additional 3 h. The colourless precipitate was vacuum-filtered through a glass filter of F-porosity and washed two times with 50 mL of methanol. The remaining 5.36 g of lithium butyryl phosphate was dried under vacuum and stored as crystalline powder at -20°C. The purity of the product was 93.6% as determined enzymatically with recombinant butyrate kinase.

2.2.2. Chemical synthesis of CoA-esters

Acetyl-CoA and butyryl-CoA were prepared and purified as previously described **[94, 95]**.

2.3. Media and plates

All media were freshly prepared and immediately autoclaved at 121°C for 20 min.

Lysogenic-Broth (LB medium), per liter

Trypton	10.0	g
Yeast extract	5.0	g
NaCl	5.0	g
Distilled water	800	mL

The medium was adjusted to pH 7.0 with NaOH (1 M) prior autoclaving.

LB-agar

Solid media were prepared by addition of the required amount of agar to the previously prepared LB-media (1.5% w/v) prior autoclaving. The temperature was adjusted to 55°C in a water-bath prior addition of antibiotic or indicator solutions.

Mineral medium (*M_9* medium)

M_9 medium consists of a *M_9* salt solution, which is complemented with compounds listed in table (2.1) prior use.

Ten-fold *M_9* salts (per liter), adjusted to pH 7.4

Na_2HPO_4	60	g
KH_2PO_4	30	g
NH_4Cl	10	g
NaCl	5	g

Table.2.1: Composition of one liter *M_9*.

Ten-fold *M_9* salts (100 mL) were diluted with 900 mL distilled water and autoclaved. The following solutions were used for completing the medium just prior use.

Component	Stock solution	volume	End Conc.	Sterilization
$MgSO_4$	1.0 M	1.0 mL	1.0 mM	Autoclaving
$CaCl_2$	0.1 M	1.0 mL	0.1 mM	Autoclaving
Thiamine-HCl. 2 H_2O	1.0 M	1.0 mL	1.0 mM	Filtration
Glucose	1.0 M	20 mL	20 mM	Filtration
$FeSO_4$	0.2 M	250 µL	50 µM	Filtration
Carbenicillin	100 mg/mL	500 µL	50 µg/ mL	Filtration
Chloramphenicol	50 mg/mL	680 µL	34µg/ mL	Filtration

2.4. Bacterial strains and plasmids used in the study

Bacterial strains and plasmids used in this study are show in table (2.2 and 2.3).

Table.2.2: Bacterial strains

Strains	Genotypes	Source
E. coli, $DH_5\alpha$-cells (Cloning strain)	F^- φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F)U169 *rec*A1 *end*A1 *hsd*R17(r_k^-, m_k^+) *pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1 λ^-	Invitrogen, Darmstadt, Germany
E. coli, Rosetta™ pLysE-cells (Expression strain)	*F^- ompT hsdS$_B$(r_B^- m_B^-) gal dcm (DE3) pLysE /RARE [argU argW ileX glyT leuW [proL] (Cam^r)*	Merck, Darmstadt, Germany

Table.2.3: Original plasmids

Plasmids	Special characteristics	Reference
pENTRY-IBA51	*Kan*^{r} : Cloning vector	IBA, Göttingen, Germany
pCFUSE-IBA11	*Amp*^{r} : Fusion vector	
pNFUSE-IBA11	*Amp*^{r} : Fusion vector	
pASG-IBAwt1	*Amp*^{r} : Expression vector	
pASG-IBA3	*Amp*^{r} : Expression vector	
pASG-IBA5	*Amp*^{r} : Expression vector	
Amp^r , Ampicillin resistance gene; *Cam*^{r} , Chloramphenicol resistance gene; *Kan*^{r} , Kanamycin resistance gene		

2.5. Oligonucleotides

Synthetic oligonucleotides were custom synthesized by MWG Biotech (Ebersberg, Germany). List of primers used in the current work are presented in the appendix.

2.6. Molecular biology kits and enzymes

GeneJET™ PCR purification kit, GeneJET™ plasmid purification kit, Phusion™ DNA Polymerase, cloning enzymes (*LguI, Esp3I,* T_4-DNA ligase) and restriction enzymes (*XbaI, HindIII, EcoRV, NlaIII, HinfI*) were purchased from Fermentas-GmbH (Fisher Scientific, Schwerte, Germany).

3. Methods

3.1. DNA Manipulations

3.1.1. PCR Amplification

Synthetic oligonucleotides were designed based on the genome-derived sequences of individual genes according to the instructions of the Stargate™ supplier (IBA, Göttingen, Germany), (Table.2.4). The *LguI* recognition site was introduced into the primers to allow the insertion of the PCR product into the cloning vector *pENTRY-IBA51*. Genomic DNA from *C. difficile* (DSM 1296T) and *C. acetobutylicum* (DSM 792T) were used as templates. Target gene amplification was performed using the PhusionTM DNA polymerase (Fisher Scientific, Schwerte, Germany). The PCR products were analyzed by either agarose gel electrophoresis or chip electrophoresis (MCE-202 MultiNA, Shimadzu Deutschland, Duisburg, Germany). Afterwards, the amplified fragments were purified with the PCR purification kit according to the manufacturer instructions.

Composition of a standard PCR mixture (100 µL final volume)

Template (10 ng/µL)	10	µL
Primer mix (10 µM each)	5.0	µL
5x HF buffer	20	µL
10 mM dNTPs	2.0	µL
Phusion™ DNA Polymerase	0.5	µL
Sterilized distilled water	62.5	µL

PCR program

Initial denaturation	98 °C	2 min	
Denaturation	98 °C	10 sec	30 cycles
Annealing	55-60 °C	20 sec	
Extension	72 °C	(30 sec/ 1kb)	
Final extension	72 °C	5 min	

3.1.2. Genes cloning

The PCR fragments were cloned into the dedicated entry vector *pENTRY-IBA51* using site-specific overhangs created by digest both the PCR fragment and the cloning vector with *LguI* and re-ligated with *T_4ligase* in one-step reaction at 30 °C for one hour.

Standard cloning reaction mixture (50 µL final volume)

PCR product	(10 ng/ µL)	10	µL
pENTRY-IBA51	(25 ng/ µL)	1.0	µL
Buffer Tango	10x	5.0	µL
DTT	(10 mM)	5.0	µL
ATP	(10 mM)	2.5	µL
LguI	(5 U/µL)	1.0	µL
T_4ligase	(5 U/µL)	1.0	µL
Sterilized distilled water		24.5	µL

The resulting ligation mixture containing *pE-GOI* (*pENTRY* carry the **G**ene **O**f **I**nterest) was transformed into chemically ultra-competent *E. coli* $DH_5\alpha$-cells **[96]**. After addition of 5 µL of ligation mixture to 70 µL of competent *E. coli* cells on ice, binding of DNA by the cells was allowed for 10 to15 min. Then, the cells were heat-shocked at 42°C for 30 sec. The cells were immediately placed back on ice and adjusted to 1 mL with LB-medium (4°C). In order to allow establishing kanamycin resistance, the cells were grown at 37°C for one hour prior plating on LB-agar with kanamycin (50 µg/mL) and X-Gal (25 µg/mL). The plates were

incubated at 37°C over night. Afterwards white colonies were picked in order to inoculate 2 mL of LB-medium with kanamycin (50 µg/mL) over night for plasmid mini-preparation.

3.1.3. Small scale plasmid DNA purification

Plasmids were isolated from 1.5 mL of an overnight culture using GeneJET™ plasmid purification kit according to the supplier instructions.

3.1.4. Determination of DNA concentration

DNA concentration was determined using a NanoDrop 1000 UV/Vis spectrophotometer (PeQlab Biotechnologie GmbH, Germany) at 260 nm against elution buffer. An absorbance of one unit equals 50 µg per mL of double-stranded DNA.

3.1.5. Restriction digests of plasmid

The plasmid identity was confirmed by restricts-fragment-length-polymorphism-(RFLP) studies with suitable endonucleases (Endonuclease digests were performed at 5 to 20 fold over-digestion). The obtained fragments were analyzed by either agarose gel or chip electrophoresis.

Standard restriction digests of plasmid (final volume of 20 µL)

Plasmid (*pE-GOI*)	(50 ng/µL)	5.0 µL
Appropriate Buffer	10x	2.0 µL
Restriction enzyme	(10 U/µL)	0.5 µL
Sterilized distilled water		12.5 µL

3.1.6. DNA agarose gel electrophoresis

PCR products and RFLP-analyses of purified plasmids were analyzed by electrophoresis in mini agarose gels (80*20*4 mm) followed by ethidium bromide staining. Therefore, low melting agarose was suspended in Tris: Acetate: EDTA (TAE) buffer and molten in a microwave. The gels were then poured into suitable casting moulds and stored after solidification at 4°C up to two weeks. After removal of the comb, the gels were placed into a horizontal electrophoresis chamber and loaded with samples completed with TriTrack DNA loading Dye (5:1 v/v). Electrophoresis was carried out at 170 V for 10 to 15 min using suitable Fast Ruler^{TM} standards for size estimation (Fisher scientifice, Schwerte, Germany). The gels were stained with ethidium bromide solution (0.5 µg/ mL) for 10 to 20 min, washed two times with deionised water and the location of DNA was visualized with UV light.

TAE buffer pH 8.4

	Stock solution (50x)	Concentration in (1x)
Tris	2.0 M	40 mM
Acetic acid	1.0 M	20 mM
EDTA	0.05 M	1 mM

3.1.7. Chip electrophoresis

In the later stages of the work, a chip electrophoresis (MCE-202 MultiNA, Shimadzu Deutschland, Dusisburg, Germany) was available in our laboratory, which almost entirely replaced gel electrophoresis in DNA fragments analysis. This system provides fully automatic analysis of multiple samples on an array of four individual electrophoresis chips for multiple uses in combination with an online detection of CyBR-Gold-mediated DNA fluorescence detection. For DNA analysis, three polyampholyte running buffer systems are available depend on fragment size (up to 500, up to 1500 and up to 3500 bp). The analysis was performed according to the manufacturer instructions.

3.1.8. DNA sequencing

Custom sequencing of DNA was performed by Eurofins MWG Operon-Biotech (Ebersberg, Germany). Standard vector-derived primers and internal primers were used for the complete sequencing of DNA templates using dye-terminator chemistry **[97]**.

3.1.9. Expression plasmid generation

Genes originally cloned in *pENTRY-IBA51* vector and confirmed by DNA sequencing were subsequently used in sub-cloning experiments (either generation of expression or fusion plasmids). Therefore, both the donor plasmid (*pE-GOI*) and the acceptor vector were digested with *Esp3I* and re-ligated with *T_4-Ligase* simultaneously at 37°C for one hour. Three different expression vectors (*pASG-IBAwt1*, *pASG-IBA3*, and *pASG-IBA5*) were used in the current study. The resulting recombinant plasmid was transformed into *E. coli* $DH_5\alpha$-cells and then plated on LB-agar plates containing carbenicillin (50 µg/ mL) and X-Gal (25 µg/ mL). The plates were incubated at 37°C over night. Afterwards white colonies were picked for plasmid mini-preparation.

Standard transfer reaction mixture (final volume of 50 µL)

Donor plasmid (*pE-GOI*)	(50 ng/ µL)	1.0 µL
Acceptor vector (expression or fusion)	(25 ng/ µL)	1.0 µL

Buffer Tango	10x	5.0	µL
DTT	(10 mM)	5.0	µL
ATP	(10 mM)	2.5	µL
Esp3I	(10 U/µL)	0.5	µL
T_4 ligase	(5 U/µL)	1.0	µL
Sterilized distilled water		34	µL

3.1.10. Fusion cloning of multiple genes

In order to construct several genes together for stepwise built-up of polycistronic operons, a fusion strategy was employed, which successively allow *Esp3I* mediated transfers of genes into dedicated fusion vectors. Two types of fusion vectors were employed (*pNFUSE-IBA11* and *pCFUSE-IBA11*), in which the first provides upstream position of the inserted GOI in the final fusion vector with a ribosomal binding site (RBS) as an intergenic region, and the second provides downstream position. Afterwards, pair-wise back-insertion of two DNA fragments individually harbored by these two vectors was assembled into *pENTRY*. Repetition of transfer into fusion vector and re-integration in the *pENTRY*-vectors can be used for a stepwise assembly of artificial operon for recombinant production of enzyme complexes (e.g. butyryl-CoA dehydrogenases) or synthetic metabolic pathways provided by several enzymatic systems.

While the transfer reaction for insertion of DNAs from *pENTRY*-vectors into fusion vectors was identical to the previously described procedure for expression vectors, the assembly of those fragments into *pENTRY* requires a slightly different protocol as it described below, but is also readily accomplished within 1 h at 30°C.

Standard transfer reaction mixture (final volume of 50 µL)

pENTRY-IBA51	(25 ng/ µL)	1.0	µL
pFF.rbs_GOI_1	(25 ng/ µL)	1.0	µL
pFFc_GOI_2	(25 ng/ µL)	1.0	µL
Buffer Tango	10x	5.0	µL
DTT	(10 mM)	5.0	µL
ATP	(10 mM)	2.5	µL
LguI	(5 U/µL)	1.0	µL
T_4ligase	(5 U/µL)	1.0	µL
Sterilized distilled water		33.5	µL

3.2. Biochemical Methods

3.2.1. Gene expression and biomass production

Gene expression was performed in *E. coli* Rosetta™-pLysE-cells harboring the expression plasmid of the individual genes. Freshly transformed cells were plated in LB-agar plates supplemented with carbenicillin (50 µg/ mL) and chloramphenicol (34 µg/ mL). After incubation overnight at 37 °C, three different colonies were picked to inoculate 2 mL LB-medium supplemented with the previously mentioned antibiotics. Over day cultures were used to inoculate 100 mL overnight cultures with the same antibiotics. Overnight cultures were used to inoculate 500 mL of the same medium and the cells were induced with anhydrotetracycline (AHT) when the optical densities reached OD_{578nm} of about 0.5. The biomass was harvested 2 h post-induction by centrifugation for 10 min at 5,000 x g. Subsequently, the pellet was washed with Tris-buffered saline (TBS), centrifuged again and stored at -80°C until used.

Composition of TBS buffer (one liter) pH= 7.2- 7.5

	Stock solution	Volume	End Conc.
Tris	1 M	10 mL	10 mM
HCl	1 M	9.2 mL	9.2 mM
NaCl	5 M	30 mL	150 mM

3.2.2. Cell free extract preparation

About one gram of frozen cells were re-suspended in 15 mL resuspension buffer (essential buffer as indicated for individual enzymes, Table 3.1) supplemented with Lysozyme (50 µg/ mL), avidin (25 µg/ mL), RNAse and DNAse (2.5 µg/ mL each). The suspension was incubated for 15 min at 37°C with gently shaking. Afterwards, it was subjected to sonication using a Branson Sonifier with a duty cycle of 50% at 60% power for 10 min in an ice-water bath. Cell debris were removed by centrifugation at 6,000 x g for 1 h at 4 °C and subsequently passed via Sephadex G25 to remove large particular matters.

Table.3.1: Essential buffer composition for individual enzymes

$THIA_1$	Tris-Cl (pH 8.0, 20 mM), DTT (1 mM), EDTA (1 mM)
BHBD	Phosphate buffer (pH 7.0, 20 mM), DTT (1 mM), EDTA (1 mM)
CRT_2	Phosphate buffer (pH 7.0, 20 mM), DTT (1 mM), EDTA (1 mM)
BCD_2	Phosphate buffer (pH 7.5, 20 mM), 5 mM $MgCl_2$, DTT (1 mM)

PBT	Phosphate buffer (pH 7.5, 20 mM), DTT (1 mM), EDTA (1 mM)
BUK	Phosphate buffer (pH 7.5, 20 mM), DTT (1 mM), 5 mM $MgCl_2$

3.2.3. Protein purification

Recombinant proteins were purified using StrepTactin™-Macroprep columns (5mL, IBA GmbH, Göttingen, Germany). Prior purification, these columns were equilibrated with at least 100 mL of the washing buffer (350 mM NaCl in essential buffer). Then, the cell-free extracts were loaded onto the column using a peristaltic pump at flow rates of about 1mL/ min. Loading of the column was followed by washing steps with 5-column volumes of washing buffer. Then, the target enzyme was eluted with elution buffer (2.5 mM desthiobiotin in washing buffer). After purification, StrepTactin-columns were regenerated with 100 mL regeneration buffer (5 mM HABA in washing buffer).

3.2.4. Determination of protein concentration

Protein concentration was determined according to the binding of Coomassie Brilliant Blue G-250 to proteins **[98]** using a Roti-Nanoquant protein assay kit (Carl Roth, GmbH, Germany). Bovine serum albumin (BSA) was used as standard. Absolute protein concentrations were determined at 280 nm and calculated based on the molar extinction coefficients of the proteins **[99]**.

3.2.5. Determination of molecular mass

The molecular mass was determined with sodium dodecylsulfate-polyacrylamide gel electrophoresis (SDS-PAGE) **[100]**. The polyacrylamide gel was casted as a separating gel topped with a stacking gel (Table 3.2). PageRuler unstained broad range protein ladder (Fermentas, Germany) was used as marker, which contains 11 bands corresponding to molecular mass of 5, 10, 15, 20, 30, 40, 50, 70, 100, 150 and 250 kDa. Cell free extract, flow through, washing and elution fractions were mixed with protein loading buffer and cooked at 95 °C for 10 min prior loading on the gel. The gel was fixed in a BioRad electrophoresis apparatus, PowerPac HC™ (USA), and the separation was performed at a constant voltage (200 V) for one hour (until the tracking dye drain off the gel). Afterwards, the proteins were stained for 30 min with hot staining solution (heating in a microwave at 380 W for 5 min). The binding dye with protein was established by drowning the gel for additional 30 min in heated fixing solution, and then the excess of the dye was taken out by de-staining the gel overnight in de-staining solution.

Table.3.2: Casting SDS-PAGE gels (all values are in µL)

Component	Final Conc.	Stacking gel		Separating gel		
		6 %	8%	10 %	12 %	15 %
Tris-HCl (pH 8.8, 1M)	375 mM	-------	-------	2,250	2,250	2,250
Tris-HCl (pH 6.8, 0.95 M)	375 mM	790	790	-------	-------	-------
Distilled water	variable	850	750	2,070	1,770	1,320
Acrlyamide/ Bis-acrylamide 40% (37/1)	variable	300	400	1,500	1,800	2,250
SDS (10% w/v)	0.1%	20	20	60	60	60
TEMED (10%)	0.05%	10	10	30	30	30
APS (10%)	0.15%	30	30	90	90	90
Final volume in µL		2,000	2,000	6,000	6,000	6,000

SDS-PAGE Solutions

- **Tris-HCl buffer (pH 8.8, 1M)**

  60.5g Tris + 152 mL HCL (1 M), then complete the volume with distilled water to 500 mL.

- **Tris-HCl buffer (pH 6.8, 0.95 M)**

  60.5g Tris + 525 mL HCL (1 M)

- **SDS (10% w/v)**

  Dissolve 20 g of SDS in distilled water and adjust the final volume to 200 mL. Then, the solution was filtered with 0.45 µm filters.

- **APS (10% w/v)**

  Dissolve 1 g of APS in 10 mL distilled water

- **TEMED (10% v/v)**

  1 mL TEMED + 9 mL distilled water

- **Electrophoresis Running Buffer (10x,** dilute 1:10 with distilled water prior use)

	Weight/ Volume	End Conc.
Tris	30 g	3.0 %
Glycine	144 g	14.4 %
SDS (10%)	100 mL	1.0 %
Distilled water	Complete the volume to one liter	

- **Protein loading Buffer (2x, 50 mL)**

	Volume		End Conc.	
Distilled water	11.5	mL		
Tris-HCl (pH 6.8, 0.95 M)	6.5	mL	125	mM
SDS (10% w/v)	20	mL	4	% (w/v)
Glycerol	10	mL	20	% (v/v)
Bromphenol blue (5%)	2	mL	0.2	% (w/v)
DTT (1M)	Add before use		20	mM

- **Staining, Fixing and De-staining solutions**

Staining solution	**Fixing solution**	**De-staining solution**
Acetic acid solution (20%) Serva Blue R (0.2% in 95% ethanol) Then mix 1:1 from the above solutions	Acetic acid (10%) Ethanol (40%)	Acetic acid (10%) Ethanol (20%)

3.2.6. Enzyme Activity measurements

Unless otherwise mentioned, all enzyme measurements were carried out at room temperature (23- 25 °C) in a final volume of 1 mL and started by the addition of enzyme.

3.2.6.1.Thiolase ($THIA_1$)

Acetyl-CoA C-acetyltransferase (thiolase, EC 2.3.1.9) activity was determined in the thiolysis direction **[21]** using CoA-SH and acetoacetyl-CoA. The decrease in the absorbance was detected at 303 nm (ε_{303nm} of 14 $mM^{-1}cm^{-1}$) **[22].** The enzyme was added to start the enzymatic reaction. A base line to account spontaneous hydrolysis of acetoacetyl-CoA in the presence of alkaline pH and magnesium ions was obtained by omitting CoA-SH. One unit of enzyme activity is defined as the amount of enzyme catalyzing the thiolytic cleavage of 1.0 µmol of acetoacetyl-CoA per min under the given conditions.

Final assay mixture for $THIA_1$ activity

Tris-HCl (pH 8)	100	mM
$MgCl_2$	10	mM
DTT	1	mM
Acetoacetyl-CoA	50	µM
CoA-SH	50	µM

$$\textbf{Unit/ mg enzyme} = \frac{(\Delta \mathbf{A}_{303nm}/\text{ min Test- } \Delta \mathbf{A}_{303nm}/\text{min Blank}) \text{ (Dilution factor)}}{14 \text{ (Enzyme volume) (mg protein/ ml Enzyme)}}$$

3.2.6.2. β-Hydroxybutyryl-CoA dehydrogenase (BHBD)

β-Hydroxybutyryl-CoA dehydrogenase (EC 1.1.1.157) activity was determined at 340 nm (ε_{340nm} of 6.22 $mM^{-1}cm^{-1}$) **[57]** by monitoring the oxidation of NADH resulting from the reduction of acetoacetyl-CoA to 3-hydroxybutyryl-CoA **[27]**. The reaction was initiated by enzyme, and the activity was calculated against a control, in which acetoacetyl-CoA was omitted. One unit of enzyme activity is defined as the amount of enzyme oxidizing 1.0 µmol of NADH per min under certain conditions.

Final assay mixture for BHBD activity

Phosphate buffer (pH 6)	50	mM
DTT	1	mM
NADH	200	µM
Acetoacetyl-CoA	50	µM

$$\textbf{Unit/ mg enzyme} = \frac{(\Delta A_{340nm}/\text{ min Test} - \Delta A_{340nm}/\text{min Blank})\ (\text{Dilution factor})}{6.22\ (\text{Enzyme volume})\ (\text{mg protein/ ml Enzyme})}$$

3.2.6.3.Crotonase (CRT_2)

3-Hydroxybutyryl-CoA dehydratase (crotonase, EC 4.2.1.55) activity was determined by monitoring the hydration of the conjugated double bond of crotonyl-CoA photometrically at 263 nm ($\Delta\varepsilon_{263nm}$ of 6.7 $mM^{-1}cm^{-1}$) **[18, 27]**. The strong dilution of the enzyme which was needed measure activity reliably was performed in TBS supplemented with BSA (1 mg/ mL). One unit of enzyme activity is defined as the amount of enzyme hydrating 1.0 µmol of crotonoyl-CoA to 3-hydroxybutyryl-CoA per min under certain conditions.

Final assay mixture for CRT_2 activity

Tris-HCl (pH 7.5)	100	mM
$MgCl_2$	5	mM
DTT	1	mM
EDTA	1	mM
BSA	50	µg
Crotonyl-CoA	50	µM

$$\textbf{Unit/ mg enzyme} = \frac{(\Delta A_{263nm}/\text{ min Test} - \Delta A_{263nm}/\text{min Blank})\ (\text{Dilution factor})}{6.7\ (\text{Enzyme volume})\ (\text{mg protein/ ml Enzyme})}$$

3.2.6.4. Butyryl-CoA dehydrogenase (BCD)

Butyryl-CoA: NAD^+ oxidoreductase (EC 1.3.99.2) activity was measured under various conditions.

3.2.6.4.1. Ferricenium hexafluorophosphate assay

Butyryl-CoA dehydrogenase activity was measured with butyryl-CoA and ferricenium hexafluorophosphate ($Fc^+FP_6^-$) as an artificial electron acceptor. The reaction of ferricenium was followed at 300 nm (ε_{300nm}= 4.3 $mM^{-1}cm^{-1}$**) [101]**. The redox indicator ferricenium hexafluorophosphate was prepared in HCl (10 mM) solution to a final concentration of 2 mM set at 617 nm (ε_{617nm} = 0.41 $mM^{-1}cm^{-1}$) **[101]**. One unit of enzyme activity is defined as the amount of enzyme reducing 2.0 µmol of ferricenium per min under certain conditions.

Final assay mixture

Potassium phosphate buffer (pH 7.5)	50	mM
Ferricenium	100	µM
Butyryl-CoA	50	µM
FAD	5	µM

$$\textbf{Unit/ mg enzyme} = \frac{(\Delta A_{300nm}/\text{ min Test} - \Delta A_{300nm}/\text{min Blank})\ (\text{Dilution factor})}{4.3 \text{ X } 2\ (\text{Enzyme volume})\ (\text{mg protein/ ml Enzyme})}$$

3.2.6.4.2. NADH oxidation under oxic condition

The rate of NADH oxidation in oxic atmosphere was determined in the absent and present of crotonyl-CoA. Transferring the electrons from NADH to O_2 (diaphorase activity) or to crotonyl-CoA (physiological activity) was followed at 340 nm (ε_{340nm} of 6.22 $mM^{-1}cm^{-1}$) **[57]**. One unit of enzyme activity is defined as the amount of enzyme oxidizing 1.0 µmol of NADH per min under certain conditions.

Final assay mixture

Tris-HCl buffer (pH 7.5)	50	mM
$MgCl_2$	10	mM
NaCl	50	mM
NADH	150	µM
Crotonyl-CoA	50	µM
FAD	5	µM

3.2.6.4.3. NADH/ crotonyl-CoA reduction under anoxic condition

The enzyme activity in the physiological direction was measured with crotonyl-CoA and NADH under anoxic atmosphere in the previously mentioned assay mixture supplemented with ferredoxin obtained from *C. pasteurianum*. The necessary reoxidation of ferredoxin was achieved with recombinant ferredoxin: NADP oxidoreductase (0.3 U/ mL) in presence of $NADP^+$ (100 µM). Enzyme activity was calculated based on monitoring the oxidation of NADH at 340 nm.

3.2.6.5. Butyrate kinase (BUK)

ATP: butyrate phosphotransferase (butyrate kinase, EC 2.7.2.7) was assayed in the butyrate-forming direction, in a coupled assay with glucokinase from *C. difficile* (EC 2.7.1.2) and glucose-6-phosphate dehydrogenase from *Bacteroides fragilis* (EC 1.1.1.49), by monitoring the increase in absorbance at 340 nm due to the reduction of $NADP^+$ (ε_{340nm}= 6.22 $mM^{-1}cm^{-1}$) **[102]**. Rate-limiting amounts of butyrate kinase were added to the assay mixture to initiate the reaction and the activity was calculated against a control, in which butyryl-phosphate was omitted. One unit of enzyme activity is defined as the amount of enzyme yielding reduction 1.0 µmol of $NADP^+$ per min under certain conditions.

Final assay mixture

Phosphate buffer (pH 7.5)	50	mM
$MgCl_2$	5	mM
NADP	200	µM
ADP	100	µM
Butyryl-phosphate	1	mM
Glucose	4	mM
Glucokinase	1	U
G-6-P dehydrogenase	1	U

$$\textbf{Unit/ mg enzyme} = \frac{(\Delta A_{340nm}/\text{ min Test} - \Delta A_{340nm}/\text{min Blank})\ (\text{Dilution factor})}{6.22\ (\text{Enzyme volume})\ (\text{mg protein/ ml Enzyme})}$$

3.2.6.6.Phosphate butyryltransferase (PBT)

Phosphate butyryltransferase (EC 2.3.1.19) activity was assayed as described above for BUK, replacing butyryl-phosphate by butyryl-CoA (50 µM) and providing BUK (1 U/ mL) in excess. The reaction was initiated by the addition of rate-limiting amounts of the enzyme and a control was done by omitting butyryl-CoA.

3.3. Butyrate production

For butyrate production under aerobic conditions, 2 mL M_9 liquid medium supplemented with carbenicillin (50 µg/ mL) and chloramphenicol (34 µg/ mL) was inoculated with one colony of *E. coli* strains harboring butyrate plasmids following incubation at 37°C with shaking at 200 rpm over day. This culture was used to inoculate 100 mL from the same medium in a 250 mL conical flask following incubation overnight. The overnight culture was used to inoculate 500 mL M_9 medium in a liter conical flask. The foreign genes were induced with AHT (200 ng/ mL) when the OD_{578} was reached of 0.5. Culture samples (1.5 mL) were taken at various time points during growth. The samples were centrifuged (13,000 rpm, 5 min), and cells and the supernatants were used in individual analyses.

3.3.1. Cell growth

Resuspended cells were used for OD_{578} determinations and subsequently recovered for enzyme tests.

3.3.2. Determination of glucose concentration

Glucose concentration was measured enzymatically based on a coupled enzyme reaction with recombinant glucokinase and glucose 6-phosphate dehydrogenase **[102]** by observing the increase of the absorbance at 365 nm due to reduction of NADP (ε_{365nm}= 3.5 $mM^{-1}cm^{-1}$) **[103]**. The standard curve was done with glucose standards and the trend-line slope was used to calculate the glucose concentration in culture dilutions.

Reaction mixture

Tris-HCl, pH 7.5	50	mM
$MgCl_2$	5	mM
DTT	2	mM
ATP	1	mM
NADP	0.2	mM
G-6-P dehydrogenase	0.2	U
Glucokinase	0.1	U

$$\textbf{Concentration (mM)} = \frac{(\Delta \mathbf{A}_{365nm}/\text{ Test - }\mathbf{A}\Delta_{365nm}/\text{ Blank})\ (\text{Dilution factor})\ \ (\text{V})\ \text{X}\ 180}{3.5\ \text{X}\ 1000\ (\text{d})\ (\text{sample volume})}$$

Where: - 180 is glucose molecular weight (g/ mol)

- 3.5 is the molar extinction coefficient at 365 nm
- (d) is the thickness (cm)
- (V) is the test volume (mL)

3.3.3. Analysis of fermentation products.

Short-chain fatty acids were analyzed by high performance liquid chromatography (HPLC) according to *Schiffels et al*.**[104]**.

3.3.3.1.Sample preparation.

Samples (300 µL) were transferred into 1.5 mL polypropylene test tubes. Two spatula of NaCl (≈ 0.2 g) and 100 µL of concentrated HCl were added and vigorously mixed with 800 µL of diethyl ether (containing 5 mM cyclohexylcarboxylic acid as an internal standard) for 1 min. The mixture was briefly centrifuged to remove residual water droplets from the ether phase. Part of the ether phase (200 µL) was transferred to 1.5 mL screw capped glass vials and mixed with a solution of oxalyl chloride (250 mM) in 200 µL of N,N-dimethyl formamide: acetonitrile (1:100 v/v) and briefly mixed. Then, 800 µL of derivatization reagent (30 mM 4-nitophenol in 500 mM pyridine/ acetonitrile) was added and mixed. The samples were directly inserted into the auto-sampler rack for measurement.

3.3.3.2.Measurement of standards and samples

20 µL from each sample was injected via auto-sampler. Binary gradients with triethylamine- acetate buffer (20 mM, pH 4.8) 35% (v/v) acetonitrile (solvent A), and the same buffer with 80% (v/v) acetonitrile (solvent B) were used for separation on a Synergi polar-RP (80 Å, 4 µm, 250 x 4.6 mm) column at 45 °C. All measurements were detected at 295 nm with a flow rate of 1 mL/ min.

3.3.4. Fermentation models

In a fermentation process, there are certain rates for compound consumption or production. These rates are usually expressed as mol/ time, e.g., glucose consumption rate in mol glucose per hour, biomass growth rate in C-mol biomass per hour, acetate and butyrate production rate in mole per hour etc. It is important to obtain the values of these rates as they occur in fermentation processes, for that the following models were used for these calculations **[105]**.

$\mathbf{Rate_X} = \frac{dM_X}{dt} = \frac{M_{X2} - M_{X1}}{t_2 - t_1} = \frac{\text{C-mmol biomass}}{\text{hour}}$	$\mathbf{Doubling\ time\ (\tau_d)} = \frac{\ln 2}{\mu^{max}} = \frac{0.693}{\mu^{max}}$
$\mu = \frac{Rate_X}{M_X} = \text{hour}^{-1}$	$\mu^{max} = \frac{\ln M_X / M_{X0}}{t} = \text{hour}^{-1}$
$\mathbf{Rate_B} = \frac{dM_B}{dt} = \frac{M_{B2} - M_{B1}}{t_2 - t_1} = \frac{\text{mmol Butyrate}}{\text{hour}}$	$q_B = \frac{Rate_B}{M_X} = \frac{\text{mmol Butyrate/ hour}}{\text{C-mol biomass}}$
$\mathbf{Rate_G} = \frac{dM_G}{dt} = \frac{M_{G2} - M_{G1}}{t_2 - t_1} = \frac{\text{mmol Glucose}}{\text{hour}}$	$q_G = \frac{Rate_G}{M_X} = \frac{\text{mmol Glucose/ hour}}{\text{C-mmol biomass}}$
$\mathbf{Yield_{(X/G)}} = \frac{Rate_X}{Rate_G} = \frac{\text{C-mmol biomass}}{\text{mmol glucose}}$	$\mathbf{Yield_{(B/G)}} = \frac{Rate_B}{Rate_G} = \frac{\text{mmol butyrate}}{\text{mmol glucose}}$
$q_G^{max} = \mu^{max}\left[\frac{M_{G(t)} - M_{G0}}{M_{X(t)} - M_{X0}}\right] = \frac{\text{mmol Glucose* } h^{-1}}{\text{C- mmol biomass}}$	

Where:

X	Biomass
dM_X	Change in total amount of biomass
$M_{X0}/M_X/M_{X1}/M_{X2}$	C-mmol of biomass at time zero/ t/ t_1/ t_2
μ	Specific growth rate
μ^{max}	Maximum specific growth rate
τ_d	Doubling time
B	Butyrate
dM_B	Change in total amount of butyrate
M_{B1}	Total amount of butyrate at time t_1
M_{B2}	Total amount of butyrate at time t_2
q_B	Specific butyrate production rate
G	Glucose
dM_G	Change in total amount of glucose
M_{G0}	mmol of glucose at zero time
M_{G1}	mmol of glucose at time t_1
M_{G2}	mmol of glucose at time t_2
q_G	Specific glucose consumption rate
q_G^{max}	Maximum Specific glucose consumption rate

4. Results

4.1. Molecular biology results

4.1.1. Background:

Butyrate formation in Clostridia is required in order to balance the intracellular redox potential upon growth with glucose. Utilization of glucose in the glycolytic pathway yields redox equivalents in form of 2 NADH per molecule of glucose. Its reoxidation is essentially required to enable further glucose consumption. In Clostridia, pyruvate as oxidation product is further oxidized to yield acetyl-CoA, carbon dioxide and reduced ferredoxin by the pyruvate: ferredoxin oxidoreductase (EC 1.2.7.1). In the NADH reoxidizing pathway, several individual enzymes are involved: The condensation of two molecules of acetyl-CoA is mediated by an acetyl-CoA C-acetyltransferase (thiolase, EC 2.3.1.9). Acetoacetyl-CoA is then reduced to 3-hydroxybutyryl-CoA by 3-hydroxybutyryl-CoA dehydrogenase (EC 1.1.1.157). From this compound, water is eliminated by the action of crotonase (EC 4.2.1.55) and the crotonyl-CoA thus formed serves as the oxidant in the butyryl-CoA dehydrogenase-(EC 1.3.99.2)-mediated oxidation of a second NADH. The latter enzyme is a complex of three subunits, the butyryl-CoA dehydrogenase and two subunits of an electro-transferring flavoprotein (ETF). The butyryl-CoA finally formed can now serve to conserve energy in form of ATP. There exist two different possibilities to achieve ATP-synthesis: The butyryl-CoA is either directly converted to butyryl-phosphate by phosphate butyryltransferase (EC 2.3.1.19), which is the substrate for ATP-regeneration by butyrate kinase (EC 2.7.2.7). In an alternative pathway, butyryl-CoA and acetate are converted to acetyl-CoA and butyrate by the butyrate CoA-transferase (EC 2.8.3.9) and acetyl-CoA is further converted via acetyl-phosphate to ATP by consecutive action of phosphate acetyltransferase (EC 2.3.1.8) and acetate kinase (EC 2.7.2.1). In a first step, the available genome of *Clostridium difficile* strain 630 was, therefore, analyzed by bioinformatics in order to establish the butyrate synthetic pathway active in this organism.

4.1.2. Identification of butyrate forming genes in *C. difficile*:

The identification of enzymes in genome-sequenced microbes is strongly facilitated by the DBGet-database (http://www.genome.jp/dbget/). Using a single EC number, this database enables the rapid screen of whole genome information for the occurrence of the necessary genes. As shown in figure (4.1), *Clostridium difficile* (and also *C. acetobutylicum*, not shown) possesses the necessary genes encoding for both pathways. From this diagram (Fig.4.1), direct links are available to the necessary data: By simple mouse-click on the EC numbers of

enzymes given in green, the amino acid sequence of the protein but also the gene sequence encoding this protein can be yielded (Fig.4.2).

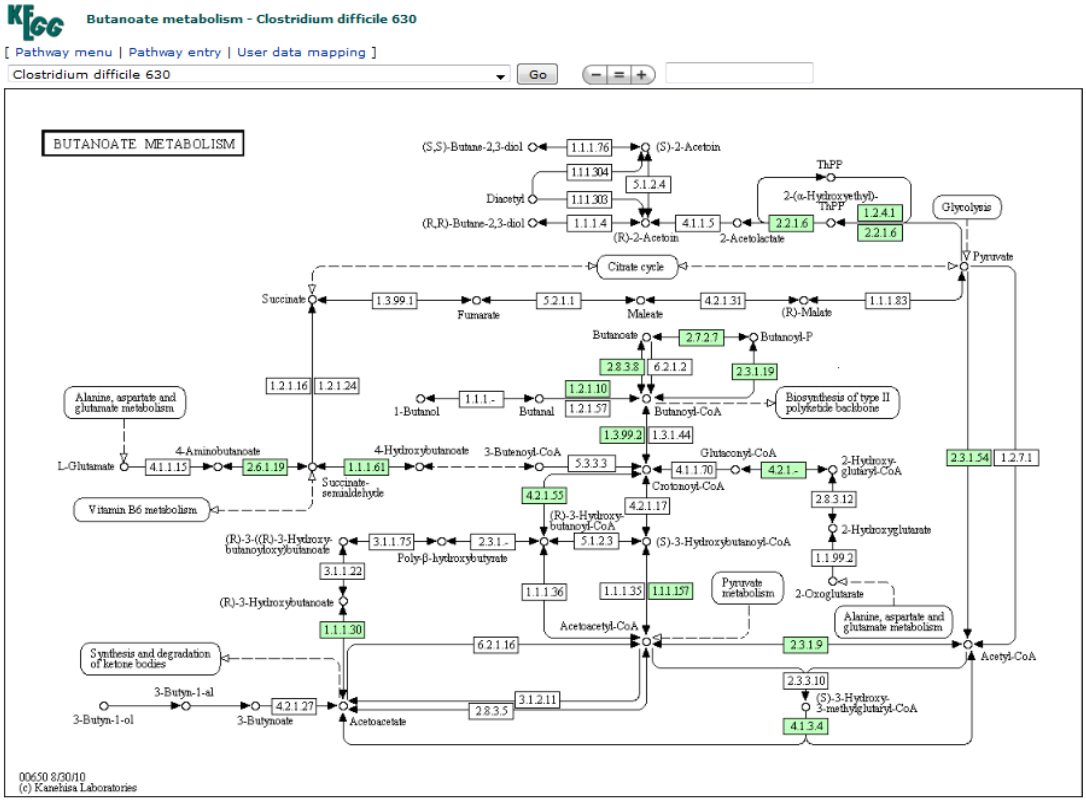


Fig.4.1: Butyrate pathway in *C. difficile* strain 630

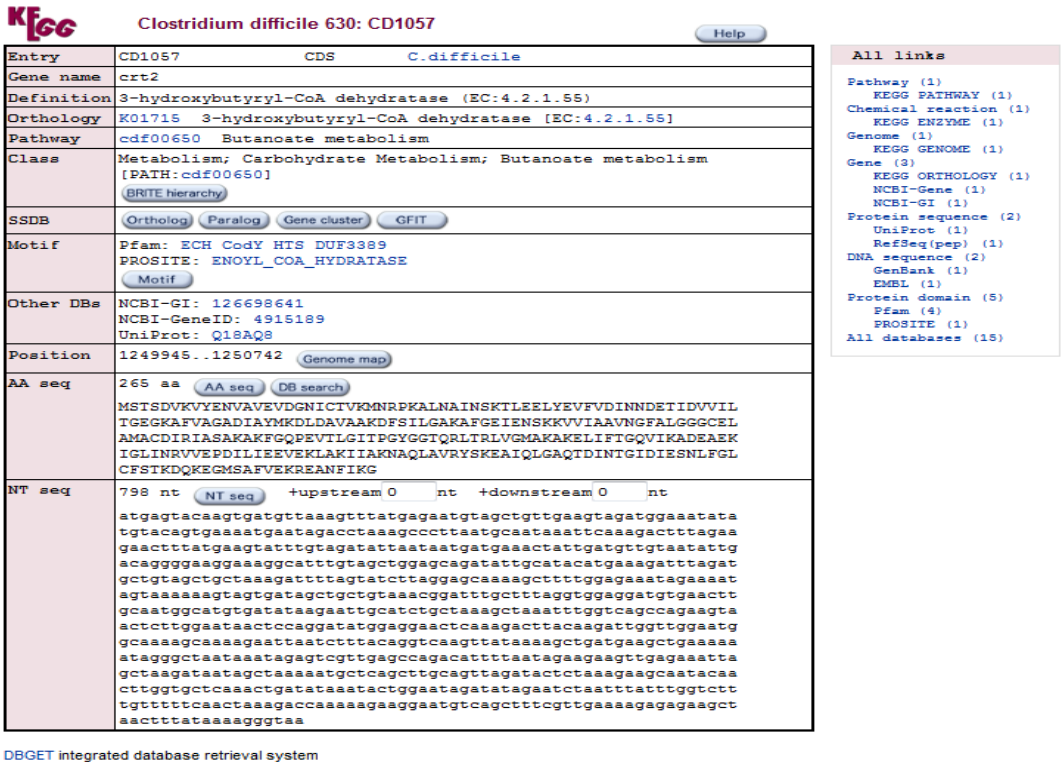


Fig.4.2: Online available data of crotonase from *C. difficile* 630

These data were first used to summarize data of the encoded proteins including amino acid content, molecular masses and isoelectric points (Table 4.1). Likewise, the corresponding genes from *C. acetobutylicum* were identified. Please note that the analysis of the butyrate CoA-transferase is actually subject of a bachelor thesis in our group and was, therefore, not considered in this thesis.

Table.4.1: Analysis of butyrate pathway biosynthetic enzymes

Name	Gene name	EC	Content (AA)	Mw (kDa)	*pI*	ε_{280}
Thiolase	*thiA*$_1$	2.3.1.9	391	40.92	5.02	0.360
3-hydroxybutyryl-CoA dehydrogenase	*hbd*	1.1.1.157	281	30.68	5.74	0.254
Crotonase	*crt*$_2$	4.2.1.55	265	28.43	4.77	0.229
Butyryl-CoA dehydrogenase	*bcd*$_2$	1.3.99.2	378	41.31	5.57	0.666
Electron transfer flavoprotein beta-subunit	*etfB*$_2$		265	28.91	4.61	0.663
Electron transfer flavoprotein alpha-subunit	*etfA*$_2$		336	36.07	4.92	0.444
Butyryl-CoA dehydrogenase	*bcd*	1.3.99.2	379	41.48	5.92	0.601
Electron transfer flavoprotein beta-subunit	*etfB*		259	28.11	5.02	0.724
Electron transfer flavoprotein alpha-subunit	*etfA*		336	28.02	5.43	0.461
Phosphate butyryltransferase	*pbt*	2.3.1.19	300	32.464	5.48	0.162
Butyrate kinase	*buk*	2.7.2.7	359	38.697	5.17	0.397

E.C, enzyme commission number **Mw, molecular weight**
ε_{280}, extinction coefficient at 280 nm (Abs 0.1% (=1 g/L) **AA, amino acid**
Highlighted rows refer to genes cloned from *C. acetobutylicum*. ***pI*, Isoelectric point**

4.1.3. Analysis of the protein-encoding genes:

Some features crucial for cloning of the enzyme-encoding genes and the recombinant enzyme production in *E. coli* might by beneficially discussed already here. In *C. difficile*, the genes encoding enzymes converting acetyl-CoA to butyryl-CoA are grouped in an operon. The average GC content of these genes is about 33.7%. The Codon Adaptation Index (CAI) of these genes as compared to the *E. coli* codon is about 0.58. The lower this number the higher is the chance that these genes will be expressed poorly **[106]**. The most problematic 15 codons are shown in figure (4.3).

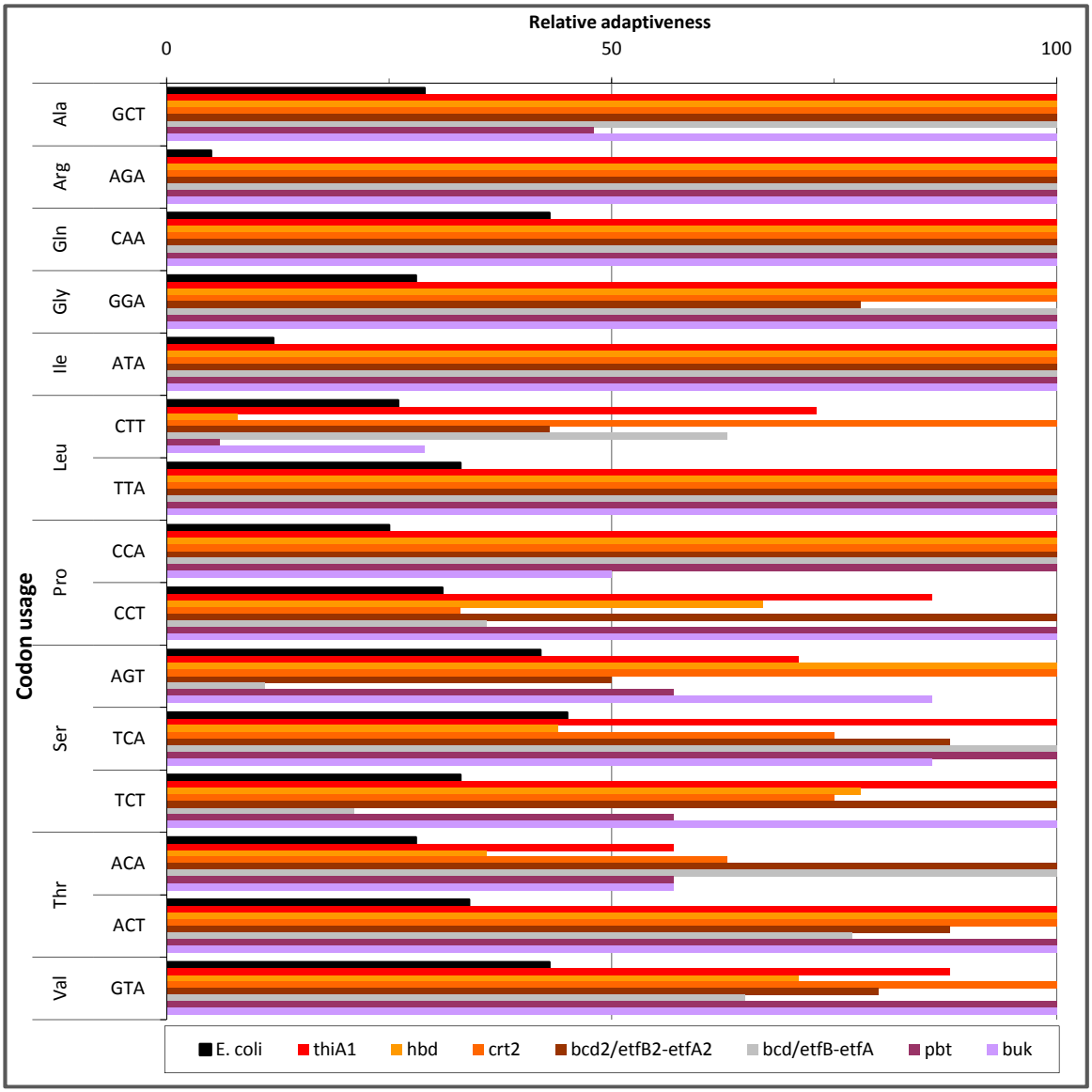


Fig.4.3: Relative adaptiveness of codons. Black bars indicate the codons usage in *E. coli.* The colored bars give the clostridial codons usage. *bcd/etfAB* are the only three genes cloned from *C. acetobutylicum*, however, the rest are cloned from *C. difficile*.

The data obtained from this analysis revealed that the clostridial codon usage differs significantly from the codon usage of *E. coli*. It should be noted, however, that the recombinant enzyme production of such proteins is frequently possible in genetically engineered strains, which provide additional copies of rare tRNA genes, and, therefore, strongly facilitate clostridial enzyme production in this host.

Features relevant for molecular cloning of individual genes like size, GC content, codon adaptation index (CAI) and the differences in codon usage between *E. coli* and clostridia are presented in table (4.2). In addition, the accession numbers are given to document the origin of genomic sequences used for planning of cloning experiments.

Table.4.2: Analysis of butyrate pathway biosynthetic genes

Gene name	Accession number	Size (bp)	GC% content	CAI	Difference in Codon Usage %
thiA$_1$	CD1059	1176	36.2	0.56	47.3
hbd	CD1058	846	29.3	0.58	48.3
crt$_2$	CD1057	798	33.0	0.59	49.6
bcd$_2$	CD1054	1137	35.1	0.59	44.9
etfB$_2$	CD1055	798	32.7	0.59	45.7
etfA$_2$	CD1056	1011	34.1	0.59	45.5
bcd	CA_C2711	1140	34.6	0.60	45.5
etfB	CA_C2710	780	35.0	0.59	45.0
etfA	CA_C2709	1011	34.7	0.57	44.9
pbt	CD0112	903	33.0	0.61	45.1
buk	CD0113	1080	32.8	0.60	46.9

The gray-highlighted rows shown *bcd, etfB* and *etfA* genes which were cloned from *C. acetobutylicum*.

4.1.4. Gene cloning and expression

In order to clone individual genes, two primers were designed (Fig.4.4) for gene amplification by PCR. The forward primer starts with the natural start codon ATG of the protein encoding genes. The sequences were chosen to provide a 100% identity with the genes and of a length ranging from 18 to 38 bp. Primer ends were usually GC-clamps of at least 2 bp. The calculated melting points of the pairing regions were chosen between 55 and 60 °C. Each sense primer then was fused with a 5'-AAGCTCTTCA-extension. This sequence contains a *LguI*-recognition site (GCTCTTC), which directs formation of a 5'-ATG-overhang in the PCR-products upon digest with this enzyme. The reverse primers for gene amplification were deduced following similar rules, including the fact that sense and antisense primers exhibited a $\Delta T_{melt} < 2$ °C. All antisense primers ended with the codon directly upstream of the natural stop codon. The antisense primer sequences were inverted and obtained a 5'-extention of the sequence AAGCTCTTCTCCC. This sequence also contains a *LguI*-recognition site, which is placed in a position to create a 5'-CCC-overhang at the PCR-product upon digest. All primers used in the current study are listed in detail in appendix (App.1).

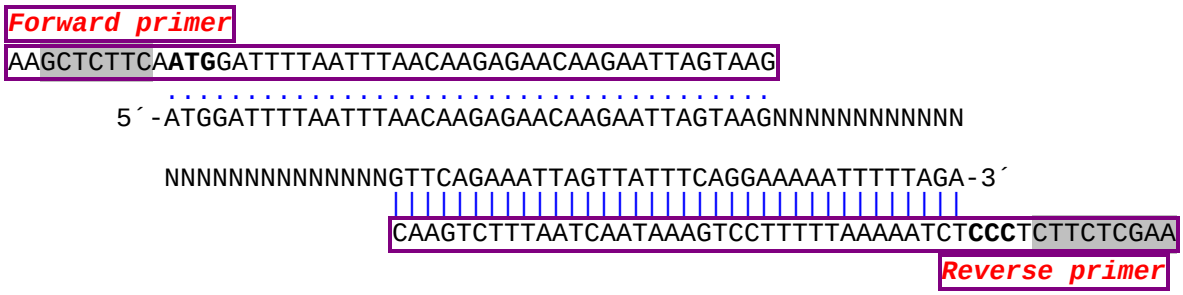


Fig.4.4: Design of forward and reverse primers. The grey highlighted sequences refer to the *LguI*-recognition site, **bold** letters show the overhang dedicated for recombination into the vector. Coding sequence is otherwise given in black, non relevant nucleotides are indicated by "N".

After analysis of the PCR products for the expected sizes by MultiNA chip electrophoresis, these products were purified and inserted into dedicated cloning vectors. The primer design allowed direct cloning of the PCR-products in vectors (*pENTRY-IBA51*), which carry a pair of *LguI* restriction sites in opposite orientation for the accommodation of the genes via the ATG- and CCC-overhang (Fig.4.5). The insertion of amplified DNA fragments into *pENTRY-IBA51* vector was achieved by simultaneously occurring digest with *LguI* and T_4-ligase mediated re-ligation. Please note that the resulting recombinant plasmids did no longer contain *LguI*-recognition sites and are therefore, resistant to *LguI*-digest (Fig.4.5).

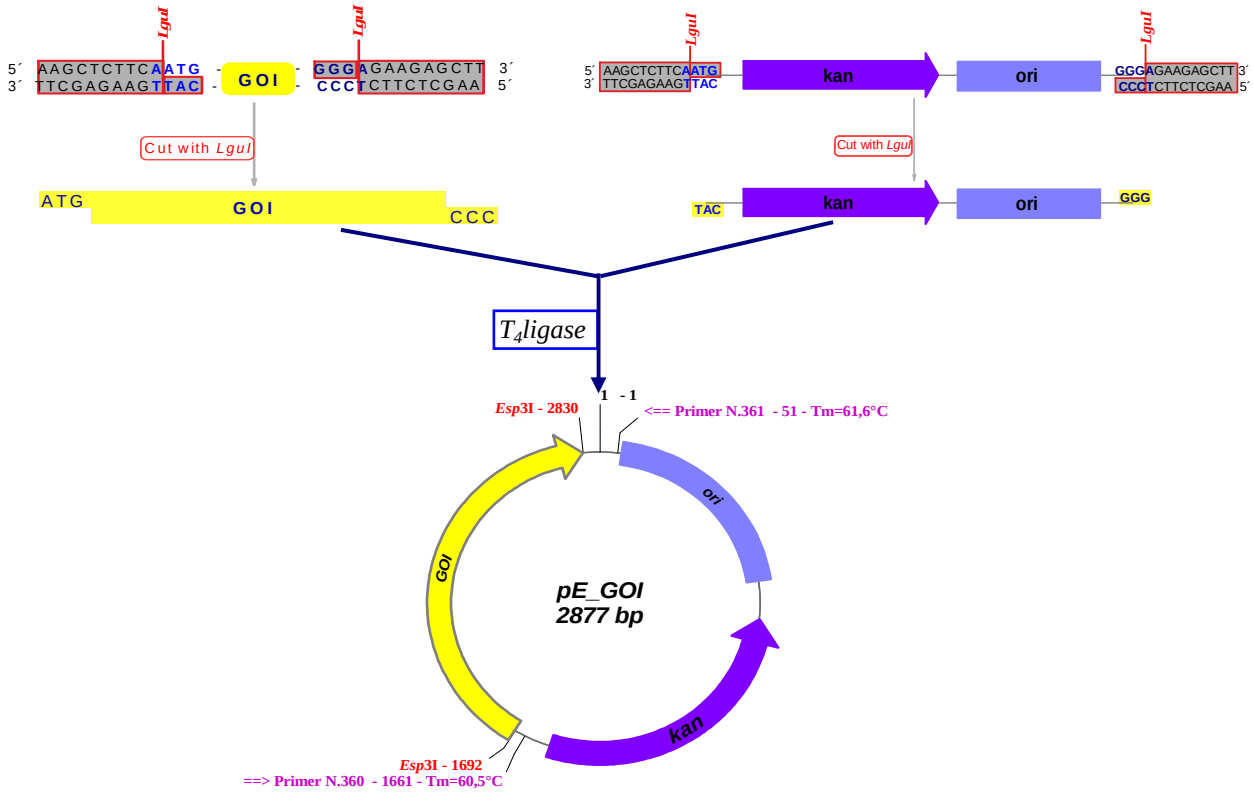
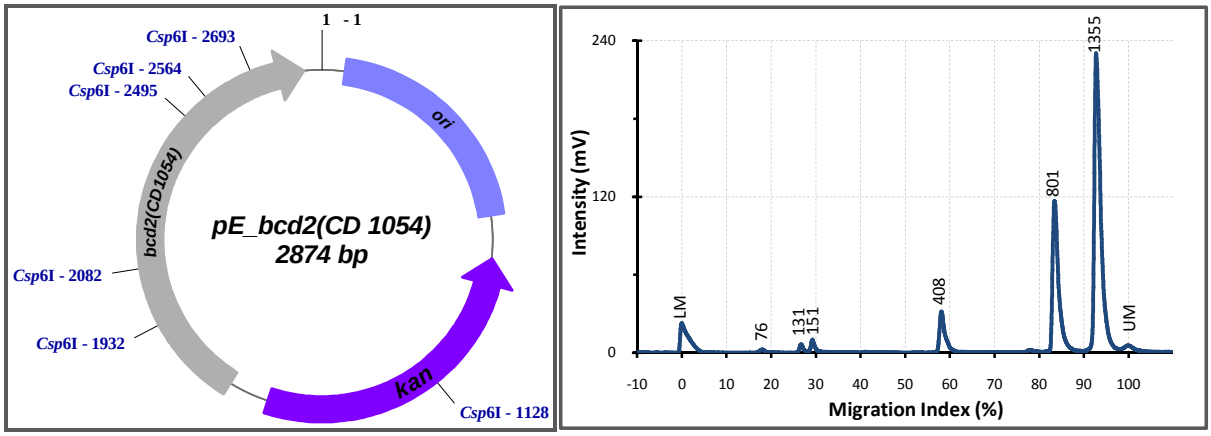


Fig.4.5: Cloning the PCR product into *pENTRY(IBA51)*. Both PCR product and *pENTRY(IBA51)* were simultaneously cut with *LguI* and ligated with T_4-ligase in one step reaction.

The highly efficient insertion of PCR products owned by this method enabled a fairly restricted choice of clones for downstream analysis. In general, 2-3 individual colonies were picked from plates obtained after transformation of *E. coli* $DH_5\alpha$ with the transfer reaction. In average, more than 90% of the resulting clones were in perfect agreement with the fragment patters predicted for the recombinant plasmids, when digested with suitable endonucleases and chip-electrophoretic analysis of the fragment sizes (Fig.4.6, more details about individual vectors analysis are presented in the supplemented appendix).



Output fragments (bp)	**LM**	76	131	151	408	801	1355	**UM**
Expected fragments (bp)	-----	69	129	150	413	804	1309	-----

Fig.4.6: Chip electrophoretic analysis of *pE_bcd$_2$ (CD1054)* fragments after suitable digest with *Csp61*. On the left hand, a pDRAW file is shown which indicatites the restriction of *Csp61*. On the right hand, an electropherogram of the digst vector is shown. LM, lower marker; UM, upper marker.

The high efficiency obtained for insertion of individual PCR-products also holds when more than one PCR-products are simultanously inserted. This feature of the cloning strategy was employed in order to create silent point mutations genes, which contained naturally occurring *LguI*- or *Esp3I*-recognition sites. Since these sequences interfere with the sub-cloning strategies further used to create multiple gene expression plasmids (see below), this silent mutation strategy at the early stage of the cloning scheme was reasonable. In order to achieve this, the naturally occurring triplet encoding an amino acid within the recognition site was replaced by an alternative codon encoding the same amino acid (e.g. a GAG glutamate codon exchanged by a GAA or vice versa). Then, this triplet was placed at the end of additional antisense and sense primers, which then allowed to use these triplets as extensions activitaed by *LguI*-recognition sites. The mutagenic triplet created an additional, unique pair of overhangs, which allowed fusion of the partially amplified genes upon insertion into the entry vector.

After thus primary testing of insertion by RFLP, the nucleotide sequences were confirmed by DNA sequencing. The sequence analysis was performed by Eurofins MWG Operon-Biotech (Ebersberg, Germany) using vector derived primers No. 360 and No.361. The positions of these primers in the plasmid sequence are indicated in figure (4.6). The sequences were then compared to the original sequences used for planning of the experiments. Since the template DNA used for cloning of the genes was derived from the wild type strain *Clostridium difficile* DSM 1296^{T} while the strain genome sequenced was strain 630 **[11]**, some minor strain-related changes were observed (Table 4.3).

Table 4.3: The observed substitutions in nucleotide and amino acid sequences.

Gene name	Accession number	Nucleotide	Amino Acid
***bcd*$_2$**	CD1054	non	non
***etfB*$_2$**	CD1055	non	non
***etfA*$_2$**	CD1056	$T_{315} \rightarrow G$ $A_{414} \rightarrow C$	non non
***crt*$_2$**	CD1057	$A_{459} \rightarrow G$ $A_{550} \rightarrow G$	non $I_{184} \rightarrow V$
hbd	CD1058	$T_{696} \rightarrow C$	non
***thiA*$_1$**	CD1059	$C_{766} \rightarrow T$ $T_{1140} \rightarrow C$	non non
pbt	CD0112	$A_{633} \rightarrow G$ $A_{706} \rightarrow C$	non non
buk	CD0113	$T_{555} \rightarrow C$ $T_{748} \rightarrow C$	non non
bcd	CA_C2711	$G_{856} \rightarrow A$	$A_{286} \rightarrow T_{286}$
etfB	CA_C2710	non	non
etfA	CA_C2709	$C_{51} \rightarrow T$	non

The shaded silent mutation in the *pbt* gene was introduced to remove the natural *LguI* recognition site. The two observed substitutions in amino acids were conserved to strong groups.

Thereafter, the cloned genes were transferred into expression vectors. While the initially generated entry clones lacked *LguI*-cleavage sites, the inserted genes were flanked by a unique pair of *Esp3I*-recognition sites. When this enzyme is used to cut the plasmids, the inserted genes are released with 5'-AATG overhangs at the coding strand and 5'-CCCT-overhangs at the non-coding strand. These overhangs are suitable for the insertion of the insert into dedicated Stargate®-pASG-Expression vectors. These vectors provide the required regulon for expression of foreign genes and a variety of vectors for the recombinant enzyme production of protein variants are available. In order to produce a wild type enzyme, the vector *pASG-IBAwt1* was used. Since the genes were originally cloned lacking the natural stop codon, this vector provides a short 3' coding extension, which fuses two amino acids

(GS) to the resulting recombinant protein and provides, thereafter, a stop codon. In order to create enzyme variants with a C-terminal StrepII-tag for affinity chromatography purification, the vector *pASG-IBA3* was used, which adds the StrepII-tag sequence GSAWSHPQFEK to the recombinant protein (Fig.4.7). The vector *pASG-IBA5* was used to create expression plasmids for production of recombinant proteins, which carry N-terminal StrepII-tags (MASAWSHPQFEKSG). This vector also provides the C-terminal extension which fuses the short GS-sequence to the C-terminus of the protein.

Since Stargate® expression vectors generally provide a unique pair of divergently oriented *Esp3I*-sites for target gene accommodation via TTAC- and GGGA-overhangs, the ATTG- and CCCT-overhangs of the released genes from the entry vector perfectly match the requirements for accommodation. Moreover, the resulting recombinant plasmid lacks *Esp3I*-recognition sites and is, therefore, resistant to further digest. Hence, transfer of the insert from the entry vector into the expression vector is easily achieved in a one-tube-reaction when *Esp3I* and T4-ligase are present simultaneously.

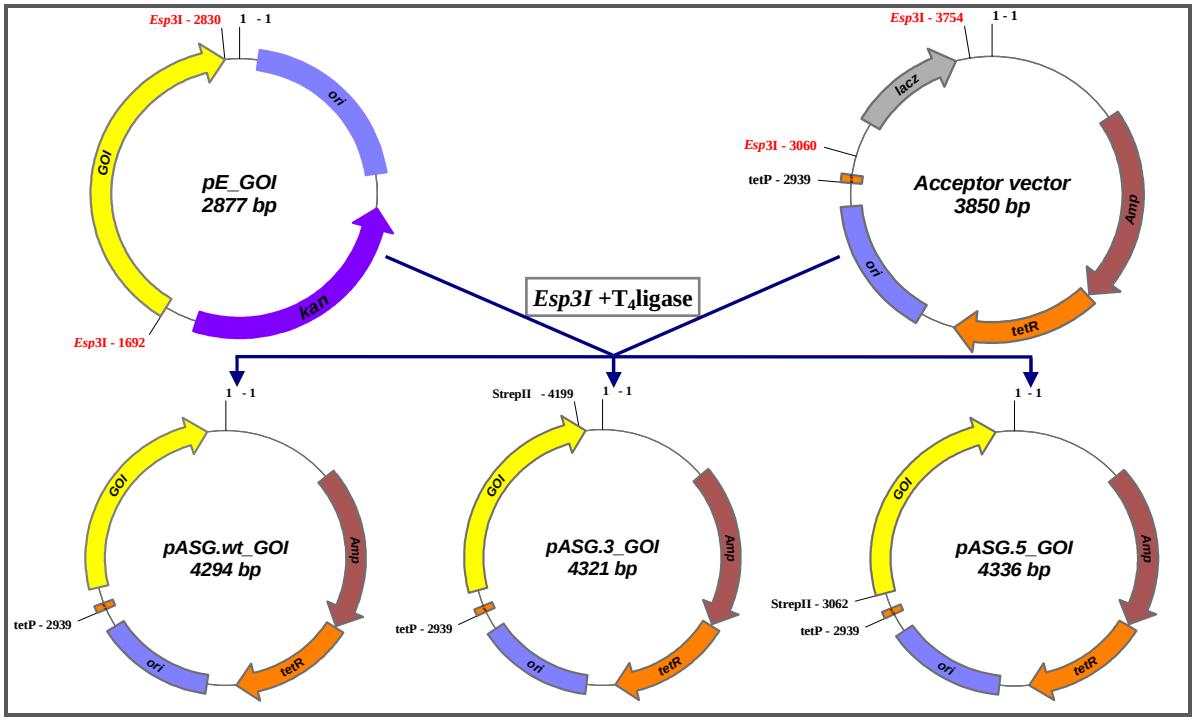


Fig.4.7: Generation of expression plasmids (*pASG_GOI*). *GOI*, gene of interest; *Amp*, ampicillin resistance gene; *tetR*, tetracycline repressor; *tetP*, tetracycline promoter; *ori*, ColE1 origin.

The very high sub-cloning efficiency (generally more than 90% of recombinant vector are formed within 1 h) and the oriented transfer of a pre-defined insert, allowed for the omission of additional sequencing of the resulting expression plasmids. Rather, correct clones were solely identified by RFLP-analysis of the prepared vector stocks and subsequent analysis by DNA chip electrophoresis.

While this strategy directly derived the expression clones for enzymes composed of identical polypeptides (e.g. acetyl-CoA C-acetyltransferase, 3-hydroxybutyryl-CoA dehydrogenase, crotonase, phosphate butyryltransferase and butyrate kinase), the butyryl-CoA dehydrogenase is a complex protein, which is composed of the dehydrogenase and two electron-transferring flavoprotein subunits. Hence, it was necessary to provide expression vectors, which allowed a balanced production of individual proteins. Moreover, it was desired to create N- and C-terminally tagged enzymes for facilitated purification. The latter was hampered by the fact that a prediction of structural interference of these tags attached to individual subunits was difficult. Thus, we decided to create 7 expression plasmids, which were either used for wild type enzyme production (note that also these proteins carried the GS-extension at their C-terminus) and variants providing N- or C-terminally StrepII-tags at either subunit.

In order to create these vectors, we took advantage of the available Stargate® fusion vectors. One of these vectors (*pNFuse-IBA11*) provides *Esp3I*-accommodation sites compatible with the donor sites of entry vectors and fuses the inserted gene with the GS-extension, a stop codon and – downstream – with an optimized ribosomal binding site for expression in *E. coli*. The second vector (*pCFuse-IBA11*) does not alter the transferred gene but creates a shift in the *LguI*-mediated overhang of the released insert (ATG to AAT). A compatible overhang (TTA) is created at the 3'-end of the assembly created within the *pNFuse*-vector by *LguI*. Taking into account that the original ATG- overhang at the 5' of the first gene and the CCC-overhang at the 3' of the second gene are retained when both fragment are released simultaneously from the fusion vectors, both inserts can be assembled and inserted into the entry vector via the TAC/GGG-accommodation sites. Obviously, a second cycle of fusion cloning starting with a bicistronic vector and the missing gene within the *pCFuse*-vector enabled the synthesis of vectors, which carried all three subunit encoding genes, in which the second and the third genes are preceded by a ribosomal binding site for efficient translation in *E. coli* (Fig.4.8). Given that three different orders of genes (ABC, BCA and CAB) were readily generated in a parallel approach, it is rather obvious that simple sub-cloning of these assemblies into dedicated expression vectors will directly yield the desired vectors for expression of various recombinant proteins, each of which is uniquely equipped with StrepII-tags at the first (*pASG-IBA5*) or the last (*pASG-IBA3*) encoded protein in the vector, while transfer into the *pASG-IBAwt1* vector yielded expression of a non-tagged enzyme variant.

It should be noted, that the assembly cloning of *C. difficile* butyryl-CoA dehydrogenase genes was performed by ***R. Fischer*** at Philipps-University, who kindly provided some vectors and expression clones used in this study. The assembly cloning of the orthologous genes from *C. acetobutylicum* are part of this work and, therefore, chosen to exemplify the cloning scheme as depicted in figure (4.8 & 4.9).

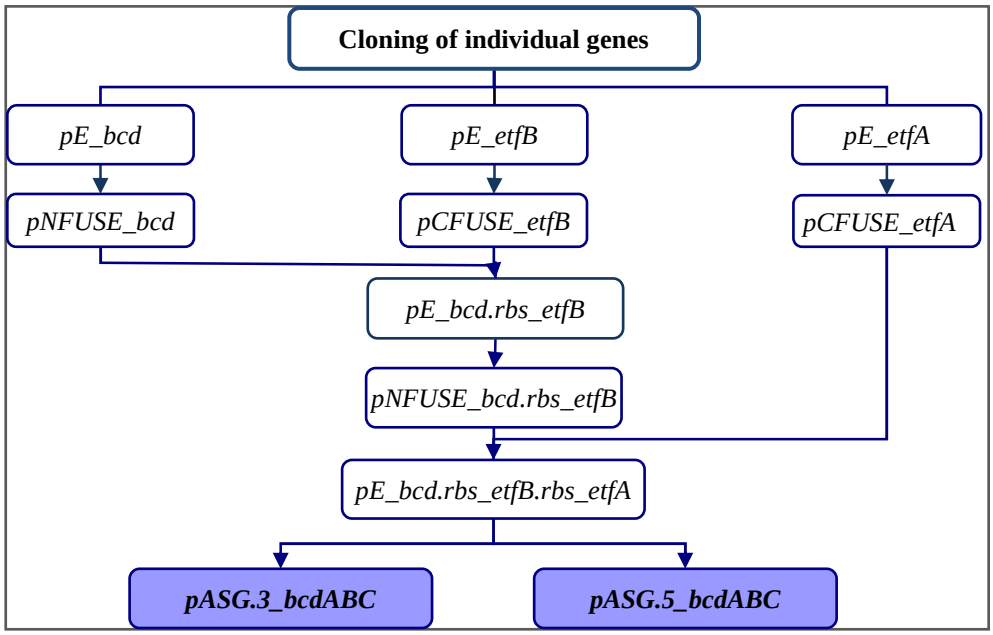


Fig.4.8: Construction of butyryl-CoA dehydrogenase and its two electron transfer subunits (*etfB* and *etfA*) from *C. acetobutylicum* as example for the fusion strategy employed in the current work.

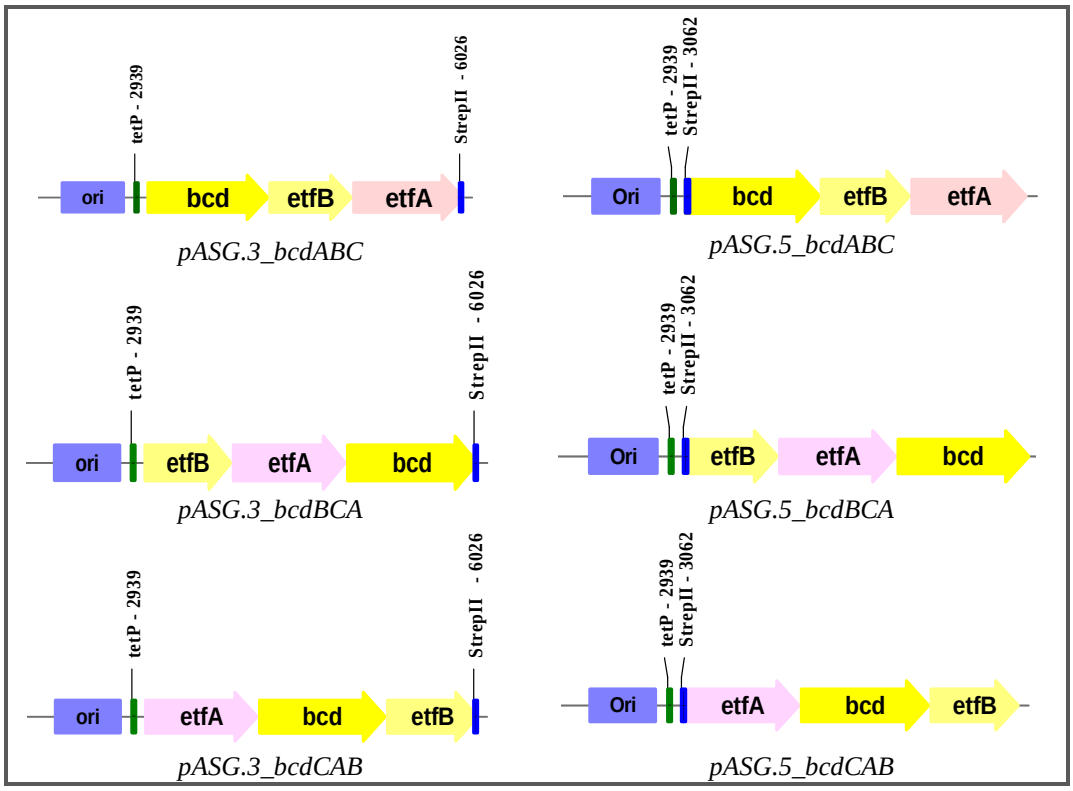


Fig.4.9: The six variants of *bcd/etf* complex from *C. acetobutylicum* constructed in the current study.

4.2. Protein purification and enzyme characterizations

The genes of butyrate synthetic enzymes from *Clostridium difficile* DSM 1296^T were expressed in *E. coli* Rosetta™ for purification and characterization. Therefore, the genes were expressed from dedicated expression plasmids whose genesis has been previously described. Specific activities, molecular masses, kinetic constants (V_{max}, K_M value, k_{cat} and k_{cat}/K_M) were determined for each of these enzymes.

4.2.1. Thiolase

Acetyl-CoA C-acetyltransferase (Thiolase, EC 2.3.1.9) is the first enzyme in the butyrogenic pathway and catalyzes the condensation of two molecules of acetyl-CoA to acetoacetyl-CoA. Thiolase from *C. difficile* has not previously been purified or characterized while the enzyme has been intensively studied from other sources. The *C*-terminally tagged enzyme was shown here as example for thiolase production and purification (Fig.4.10) because it was purer than *N*-terminally tagged enzyme or the wild type. The protein was purified close to apparent homogeneity as judge by 12% SDS-PAGE after Coomassie staining (Fig.4.10) showing a single band corresponding to a molecular mass of 45.1 kDa which is in agreement with the calculated molecular mass (about 42 kDa) using Biology Workbench (version 3.2). Thiolase behaves as an apparent tetramer with a native molecular mass of 180±5 kDa (4*45 kDa) in native gel electrophoresis (Fig.4.10).

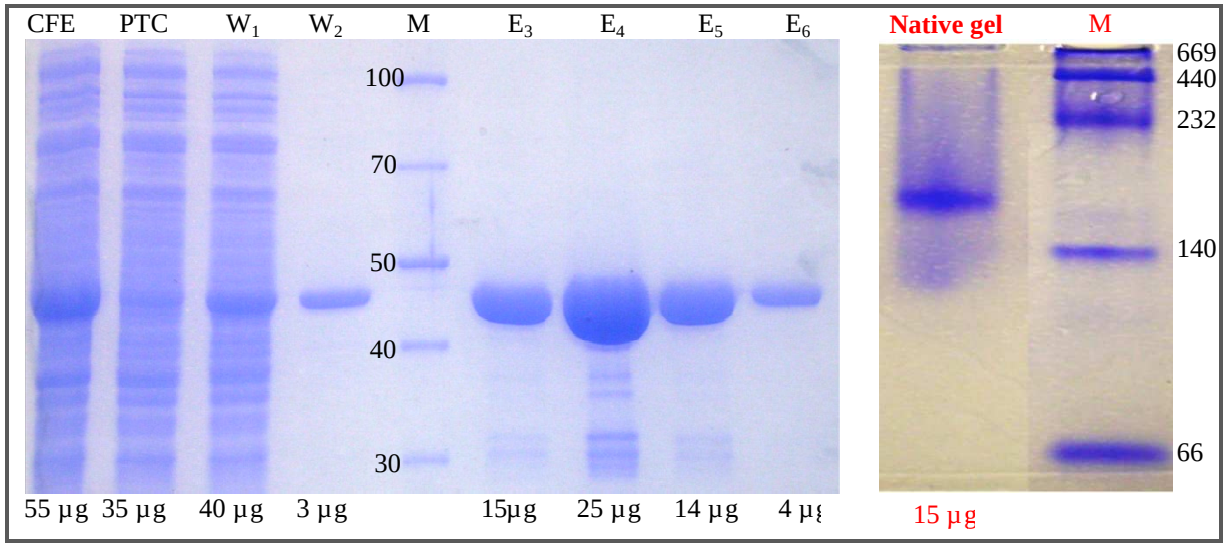
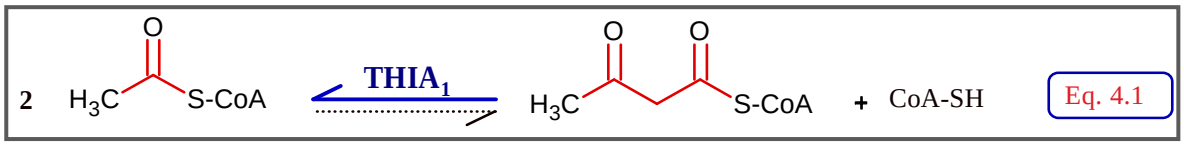


Fig.4.10:Purification of *C*-terminally tagged Thiolase. The loading protein amount is written below each lane.

The left gel is 12% SDS-PAGE Stained with Coomassie. CFE, cell free extract; PTC, pass through the column; W, wash fractions; E, elution fractions; M, PageRuler™ Unstained Protein Marker.

The right gel is 7% Native-PAGE showing the native molecular mass for Thiolase using the following protein markers: Thyroglobulin, 669 kDa; Ferritin, 440 kDa; Catalase, 232 kDa; Lactate dehydrogenase, 140 kDa; Bovine serum albumin, 66 kDa.



The enzyme activity was measured in the thiolytic direction (Eq. 4.1; bold blue arrow) by monitoring the decrease in absorbance at 303 nm at room temperature using a molar extinction coefficient of 14 $mM^{-1}cm^{-1}$ **[22]**. The assay contained Tris-HCl (pH 8, 100 mM), $MgCl_2$ (10 mM), DTT (1 mM), acetoacetyl-CoA and CoA-SH (50 µM each). The specific activity was about 29 U/mg in cell free extract (CFE) and increased to 532 U/mg for the pure enzyme with 18 fold-purified proteins.

Kinetic Constants: The kinetic constants were determined in the thiolytic direction (Eq. 4.1), under conditions previously described, with varying acetoacetyl-CoA concentrations (5- 100 µM) in the presence of fixed concentrations of CoA-SH (1.5- 100 µM) and the pure enzyme (21 ng) per test. Lineweaver-Burk plots were obtained by plotting reciprocal initial thiolysis velocity versus reciprocal CoA-SH concentration in the presence of fixed concentrations of acetoacetyl-CoA (Fig.4.11). The parallel straight lines indicate that the kinetic mechanism for thiolase is a Ping Pong pattern involving an acetyl-enzyme covalent intermediate. In order to calculate the apparent K_M and V_{max} values graphically, two different secondary graphs were plotted using the data from the primary graph (Fig.4.11). In the first graph (Fig.4.12.*A*), the linear reciprocal plot of varied acetoacetyl-CoA concentrations was plotted against ordinate intercepts which were obtained from figure (4.11). The apparent K_M value for acetoacetyl-CoA was about 33 µM with an estimated V_{max} of 625 U/mg. A second graph (Fig.4.12.*B*) was employed to calculate the apparent K_M for CoA-SH, in which the linear reciprocal plot of varied acetoacetyl-CoA concentrations was plotted versus abscissa intercepts which were derived from figure (4.11). As a result, thiolase had an apparent K_M about 27 µM for CoA-SH. The inhibitor effect of CoA-SH was observed at low concentration of acetoacetyl-CoA and high concentration of CoA-SH (Fig.4.11). The k_{cat} value in the thiolytic direction was 470 s^{-1} and k_{cat}/K_M was 14 $\mu M^{-1}s^{-1}$.

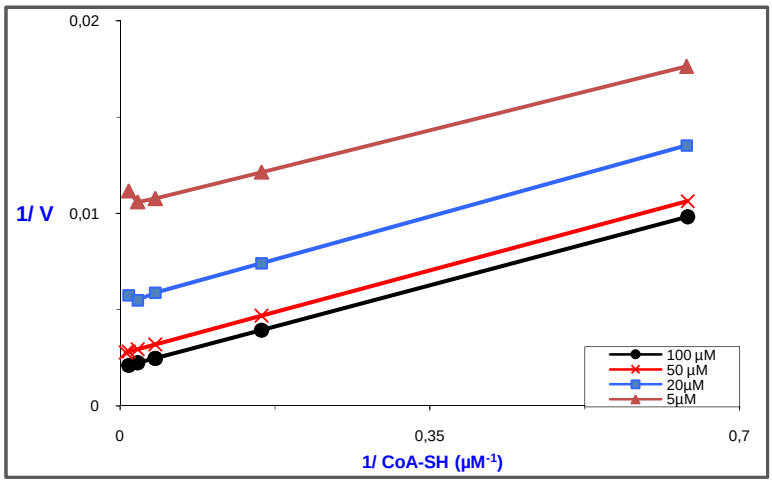


Fig.4.11: Double reciprocal plots of thiolase activity at varied CoA-SH concentrations in the presence of fixed concentrations of acetoacetyl-CoA as indicated in legend.

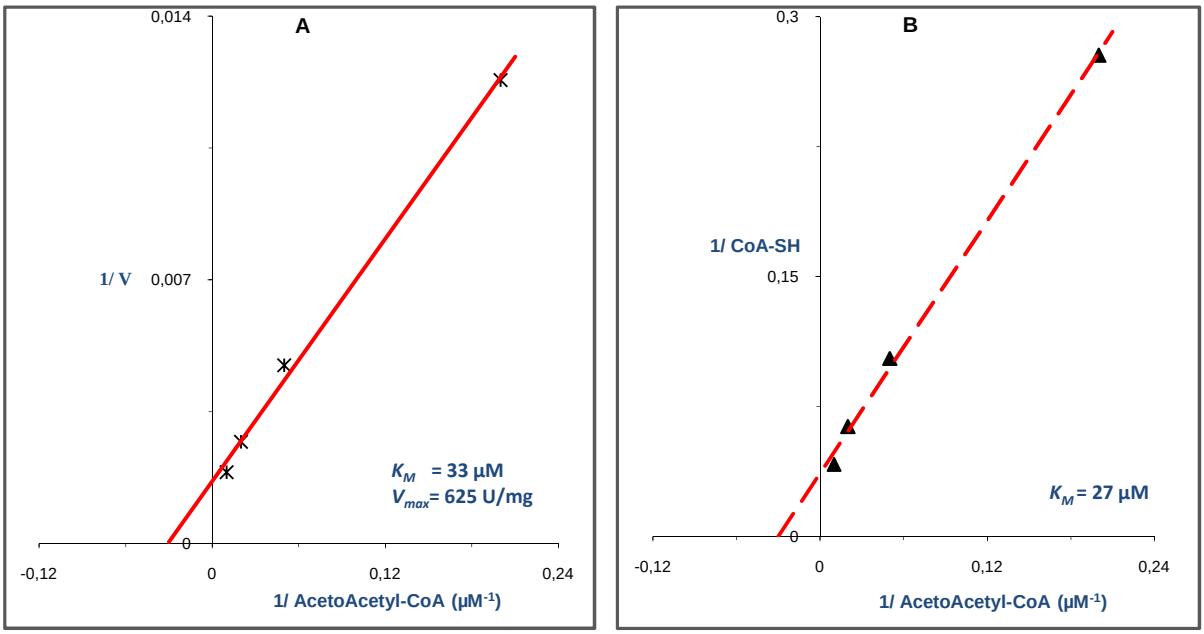


Fig.4.12: Secondary plots of reciprocal acetoacetyl-CoA concentration against ordinate intercepts (*A*) and abscissa intercepts (*B*) which were obtained from figure (4.11).

4.2.2. β-Hydroxybutyryl-CoA dehydrogenase

β-Hydroxybutyryl-CoA dehydrogenase (BHBD, EC 1.1.1.157) catalyzes the reduction of acetoacetyl-CoA with NADH to form 3-hydroxybutyryl-CoA. This subprocess is the initial and necessary step towards the ultimate production of butyrate and butanol. The *hbd* gene, encoding BHBD from *C. acetobutylicum* ATCC 824 and *C. difficile*, has already been cloned in *E. coli* and sequenced **[25, 26]**, but it had not yet been purified or characterized from *C. difficile*. This work describes the purification to homogeneity of BHBD from *C. difficile* as it is shown in 15% SDS-PAGE for the *C*-terminally tagged BHBD with a monomeric molecular mass of 33.5 kDa, correlating with the predicted molecular mass (32 kDa). The enzyme was determined to be a tetramer with a native molecular mass of 144±3 kDa (4*33.5) (Fig.4.13).

The enzyme activity was measured in the physiological direction (reduction of acetoacetyl-CoA, Eq. 4.2) by monitoring the decrease in absorbance at 340 nm at room temperature using a molar extinction coefficient of 6.22 $mM^{-1}cm^{-1}$ **[57]**. The assay contained Kpp (pH 6, 50 mM), DTT (1 mM), acetoacetyl-CoA (50 µM) and NADH (200 µM). The specific activity in the CFE was 1.3 and 1.1 U/mg for the *C*- and *N*-terminal tagged enzyme, respectively. The purification proceeded 20 and 18 fold with specific activities of 26 and 20 U/mg, respectively. The BHBD from *C. difficile* showed a relative activity about 3% with NADPH compared to NADH indicating that BHBD is NADH-depended.

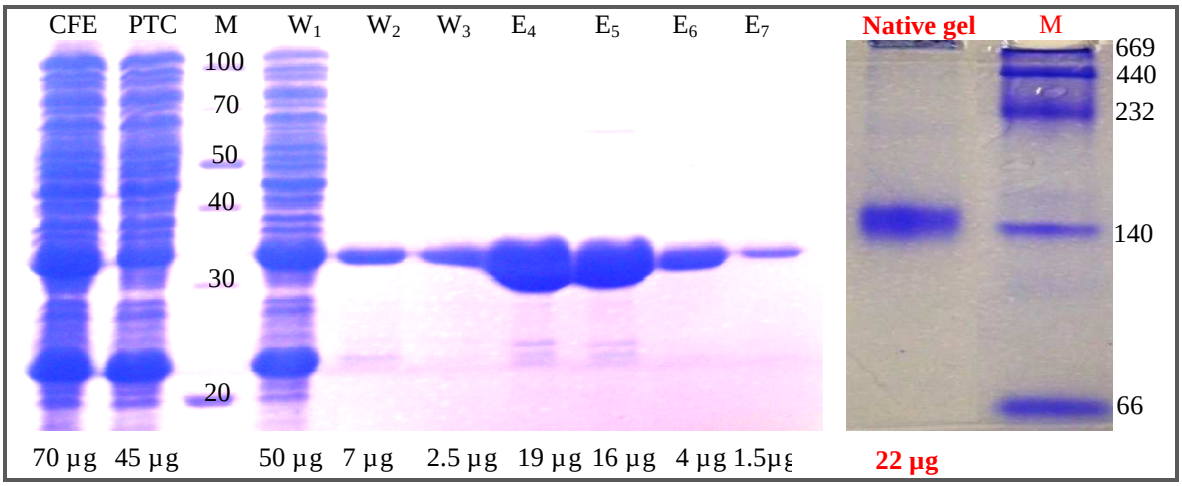
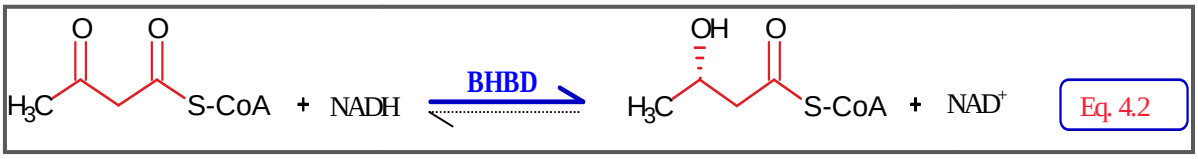


Fig.4.13: Purification of *C*-terminally tagged BHBD. The loading protein amount is written below each lane.

The left gel is 15% SDS-PAGE Stained with Coomassie. CFE, cell free extract; PTC, pass through the column; W, wash fraction; E, elution fractions; M, PageRuler™ Unstained Protein Marker.

The right gel is 7% Native-PAGE showing the native molecular mass for BHBD using the following protein markers: Thyroglobulin, 669 kDa; Ferritin, 440 kDa; Catalase, 232 kDa; Lactate dehydrogenase, 140 kDa; Bovine serum albumin, 66 kDa.



Kinetic Constants. Kinetics properties for BHBD were analyzed in the acetoacetyl-CoA reduction direction (Eq. 4.2) by variation of acetoacetyl-CoA concentrations (3- 100 µM) in the presence of fixed concentrations of NADH (25- 400 µM) using 0.9 µg pure *C*-terminally tagged enzyme under the previously described conditions. Figure (4.14) shows that this enzyme exhibits substrate inhibition and the pattern of inhibition is related to the saturation with NADH. In order to calculate K_M and V_{max} values graphically, Lineweaver-Burk plots were employed. In the primary plot (Fig.4.15), the initial velocity at various NADH concentrations is plotted reciprocally versus acetoacetyl-CoA concentrations. Obviously, the enzyme is inhibited by acetoacetyl-CoA. Consequently, the inhibitor effect at high acetoacetyl-CoA concentration was ignored to calculate the apparent K_M (Fig.4.15), which was determined to be 17 µM for acetoacetyl-CoA. In addition, plotting NADH concentrations against the ordinate intercepts that were obtained from figure (4.15) derived the apparent K_M for NADH, calculated as 210 µM with an apparent V_{max} of 54 U/mg (Fig.4.16). The calculated k_{cat} value for BHBD in the redaction direction of acetoacetyl-CoA was 30 s^{-1} and K_{cat}/K_M was 1.8 $\mu M^{-1}s^{-1}$.

Fig.4.14: Michaelis-Menten plots of BHBD activity.

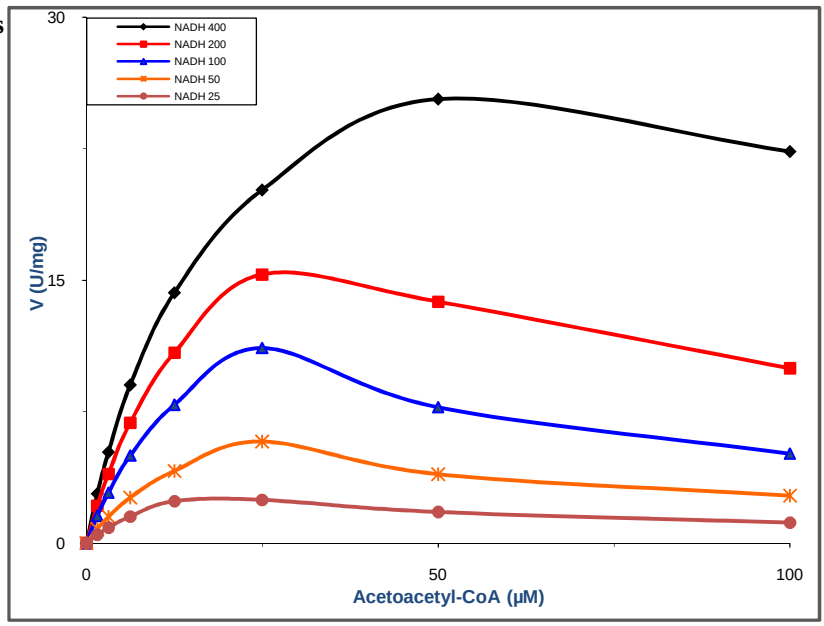


Fig.4.15: Double reciprocal plots of BHBD activity.

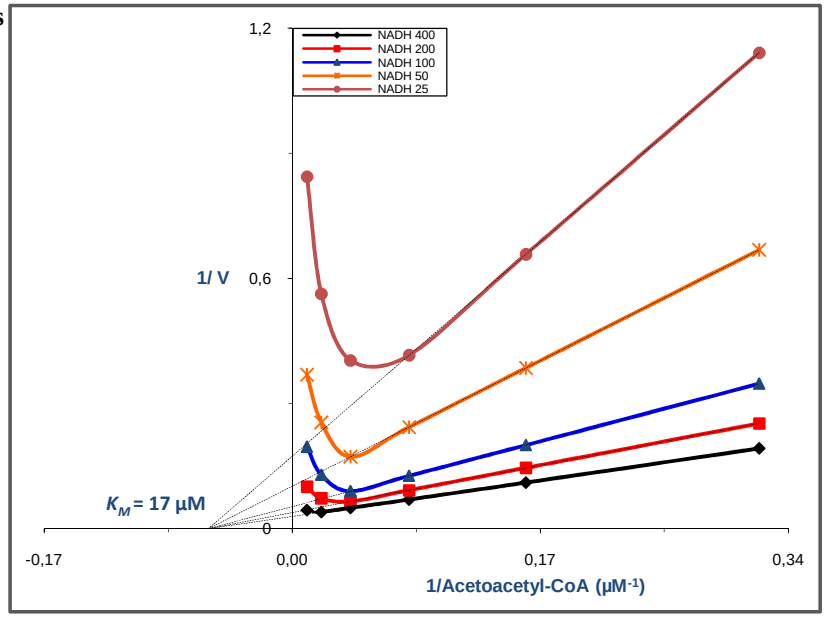
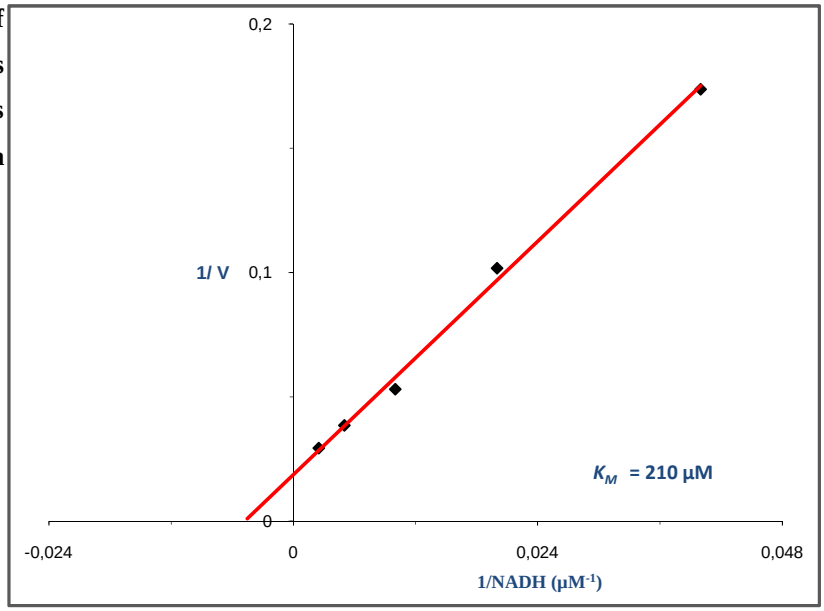


Fig.4.16: Secondary plots of NADH concentrations versus ordinate intercepts that were obtained from figure (4.15).



The k_i value for acetoacetyl-CoA was obtained from Dixon plots, in which acetoacetyl-CoA concentration was plotted against reciprocal velocity in two different graphs, either in the presence of un-saturated fixed concentrations of NADH (25- 100 µM) (Fig.4.17.*A*), or in the presence of fixed saturating concentrations of NADH (200 and 400 µM) (Fig.4.17.*B*). The data show that acetoacetyl-CoA was a non-competitive inhibitor for BHBD with an apparent k_i of 43 µM.

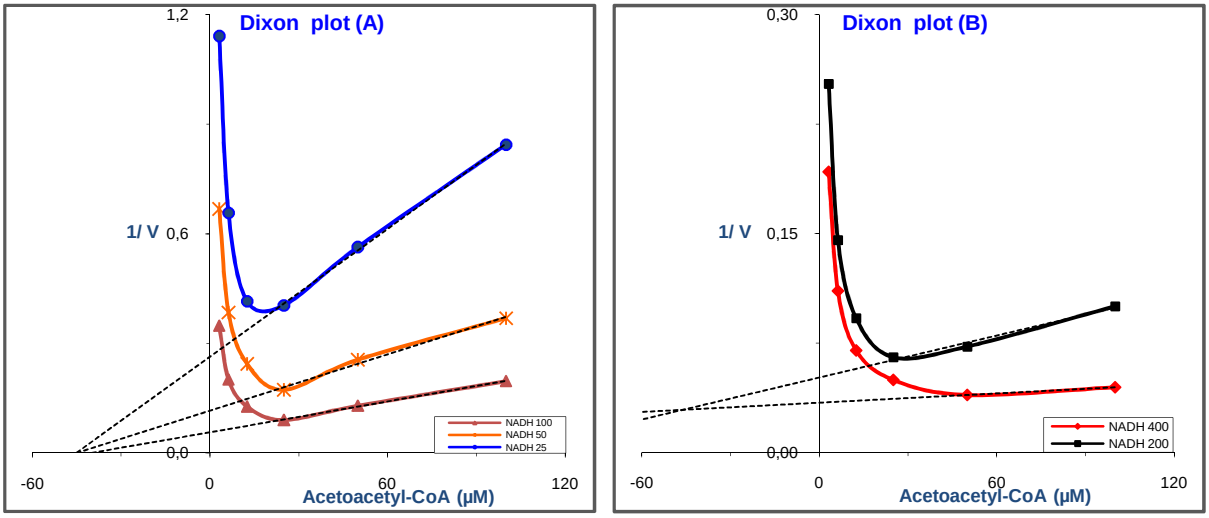


Fig.4.17. Dixon plots for BHBD at different concentrations of acetoacetyl-CoA in the presence of non-saturated (*A*) or saturated (*B*) fixed NADH concentrations.

Mechanistically, the data shown indicate a non-competitive inhibition in the presence of un-saturated concentrations of acetoacetyl-CoA (Fig.4.18.*A*). On the contrary, competitive inhibition was observed in the presence of saturated concentrations of acetoacetyl-CoA (25-100 µM) (Fig.4.18.*B*). These findings are compatible with a steady-state ordered *bi-bi* sequential bi-reactant kinetic mechanism with NADH binding first to the free enzyme.

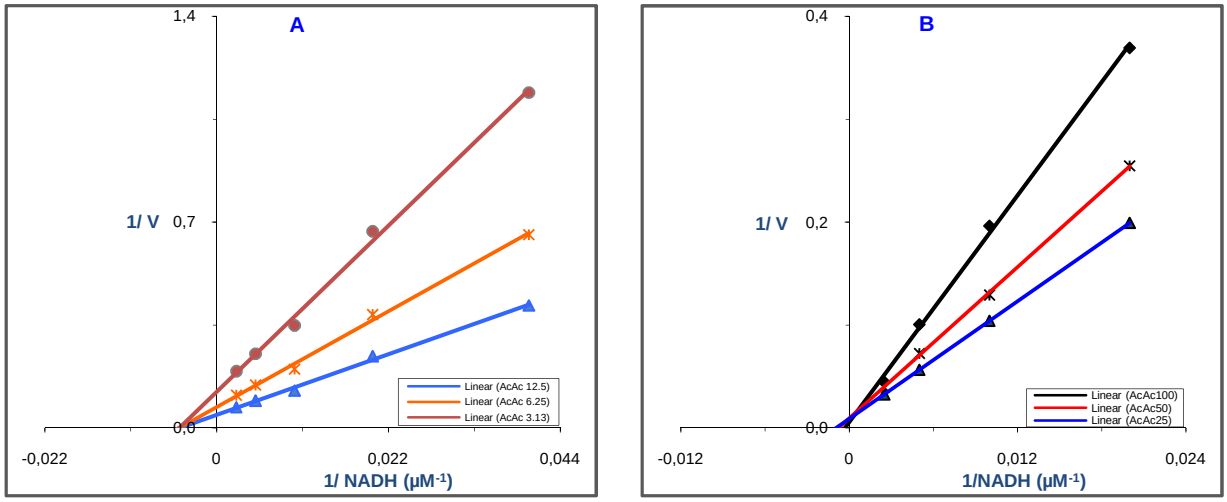
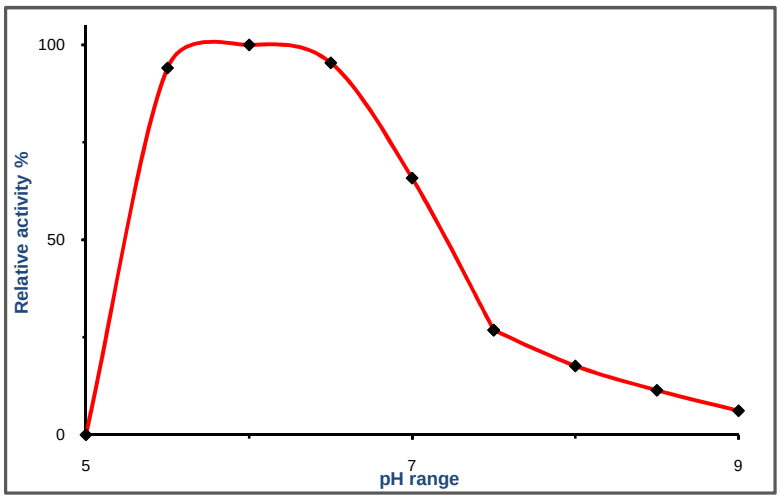


Fig.4.18. Double reciprocal plots for BHBD activity in the presence of un-saturated (*A*) or saturated (*B*) fixed acetoacetyl-CoA concentrations.

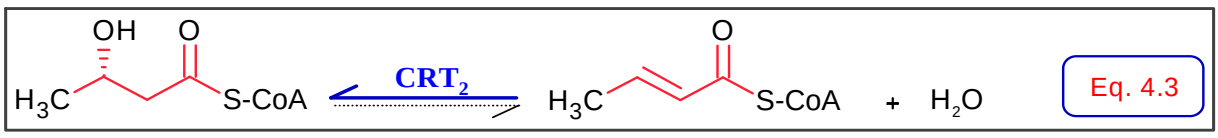
pH optimum: Activity for BHBD was measured at different pH values using acetate buffer (pH 5.5 and 6.0), phosphate buffer (pH 6.0- 7.5) and Tris-HCl buffer (pH 7.5- 9.0). Noteworthy, the enzyme precipitated under acidic conditions (acetate buffer pH less than 5.5). The enzyme showed high relative activities from pH 5.5 to 6.5 (Fig.4.19) with an optimum pH of about 6. When the enzyme works in neutural or light alkaline conditions, the relative activity sharply decreases to 67 and 27% in pH 7.0 and 7.5, respectively.

Fig.4.19: Optimum pH for BHBD.



4.2.3. Crotonase

3-Hydroxybutyryl-CoA dehydratase (Crotonase, EC 4.2.1.55) capable of catalyzing the reversible hydration of crotonyl-CoA to hydroxybutyryl-CoA from *C. difficile* has been purified to apparent homogeneity. The purity of *C*-terminally tagged crotonase was confirmed via 15% SDS-PAGE (Fig.4.20) showing a single band corresponding to a molecular mass of 29.7 kDa, close to the calculated molecular mass. The native molecular mass was 125±3 kDa (Fig.4.20), suggesting that it is a tetramer (4*29.7). The enzyme activity was measured by monitoring the decrease of crotonyl-CoA due to hydration of the conjugated double bond at 263 nm (the reverse direction of Eq. 4.3) **[27]** at room temperature using a molar extinction coefficient of 6.7 $mM^{-1}cm^{-1}$ **[18, 27].** The assay contained Tris-HCl (pH 7.5, 100 mM), $MgCl_2$ (5 mM), DTT (1 mM), EDTA (1 mM), Bovine serum albumin (50 µg) and crotonyl-CoA (50 µM). The enzyme was significantly more stable when diluted in a BSA containg buffer (1 mg/mL) as compared to dilution in buffer without BSA. The specific activity in cell extracts was 705 U/mg for *C*-terminally tagged enzyme. The enzyme could be purified 26 fold with specific activity of 18,600 U/mg.



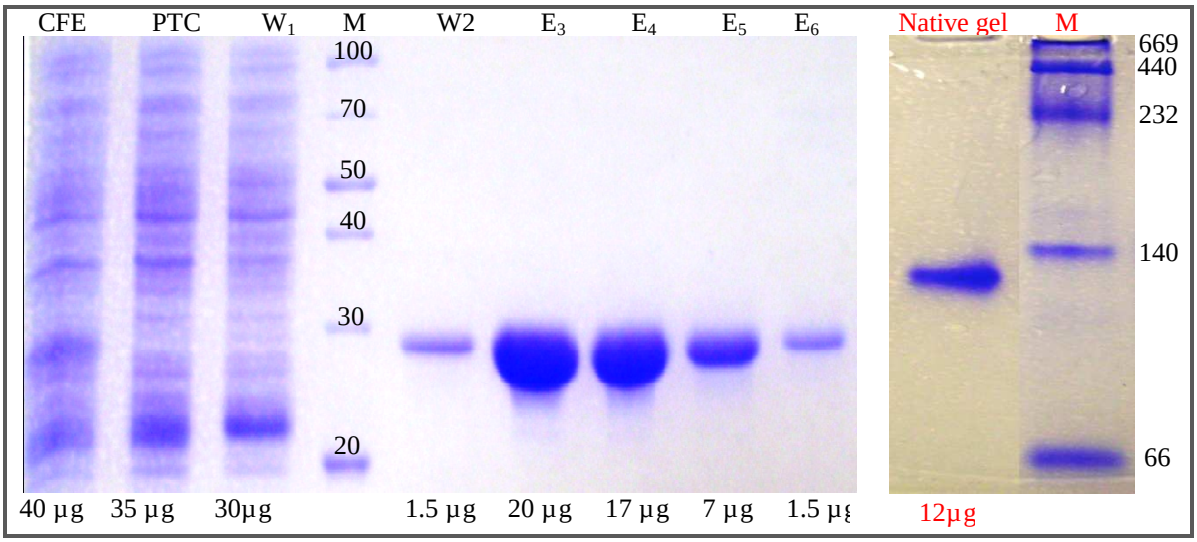


Fig.4.20: Purification of *C*-terminally tagged CRT_2. The loading protein amount is written below each lane.

The left gel is 15% SDS-PAGE Stained with Coomassie. CFE, cell free extract; PTC, pass through the column; W, wash fraction; E, elution fractions; M, PageRuler™ Unstained Protein Marker.

The right gel is 7% Native-PAGE showing the native molecular mass for CRT_2 using the following protein markers: Thyroglobulin, 669 kDa; Ferritin, 440 kDa; Catalase, 232 kDa; Lactate dehydrogenase, 140 kDa; Bovine serum albumin, 66 kDa.

Kinetic Constants. *C*-terminally tagged crotonase was employed to calculate apparent V_{max} and K_M values because the purity and yield were better than the *N*-terminally tagged one. Values for V_{max} and K_M were calculated from Lineweaver-Burk plots (Fig.4.21) as well as from non-linear regression of Michaels-Menten plots (Fig.4.22) in the hydration direction (reverse direction of Eq. 4.3) employing previously described conditions with 1.7 ng enzyme and varying concentrations of crotonyl-CoA (3- 150 µM). The apparent K_M for crotonyl-CoA was about 52 µM with a V_{max} of $22.5*10^3$ U/mg. The calculated turnover number (k_{cat}) for crotonase was about $11.12*10^3$ s^{-1} and K_{cat}/K_M was about 214 $\mu M^{-1}s^{-1}$.

Fig.4.21: Double reciprocal plots of crotonase.

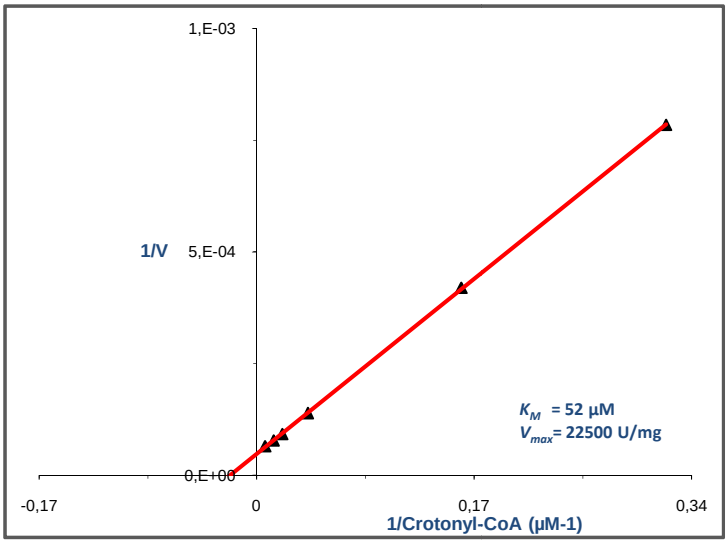
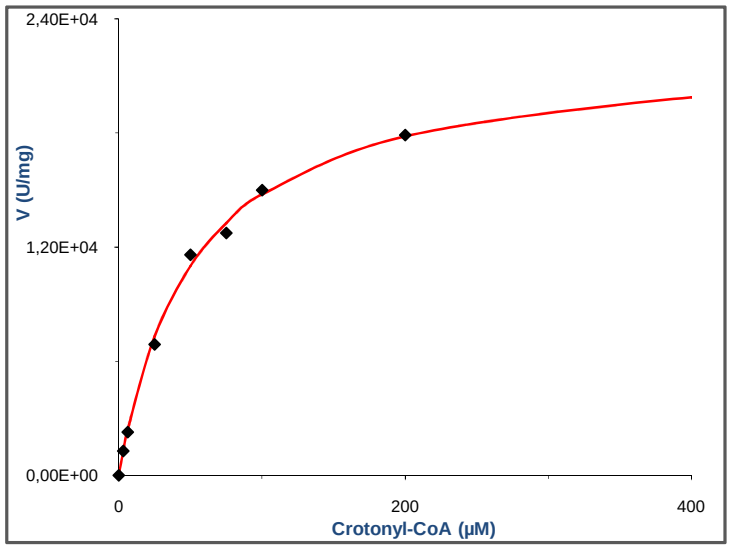


Fig.4.22: Michaelis-Menten curve for Crotonase.



4.2.4. Butyryl-CoA dehydrogenase/ ETF complex

As stated initially, one of the most important aims of this study was to figure out the nature of butyryl-CoA dehydrogenase from *Clostridium difficile*. The finding that the genome of *C. difficile* harbors *rnf*-genes, which could allow energy conservation aided by reduced ferredoxin produced by a bi-forking enoyl-CoA reductase, let us to expect such an enzyme in this organism. In contrast, the genome of *C. acetobutylicum* lacks *rnf*-genes, suggesting that a common butyryl-CoA dehydrogenase is present in the latter organism. Thus, the enzyme from *C. acetobutylicum* was introduced in this study for control purposes.

The previous work on butyryl-CoA dehydrogenase/ETF complex (EC 1.3.99.2) constructed from *C. difficile* showed that the six enzyme variants using the different constructed plasmids (Fig.4.9) expressed in *E. coli* have a different activity behavior in cell free extract than the pure proteins **[107]**. These investigations have shown that only an enzyme variant which carries a *C*-terminal StrepII-tag at the etfA-subunit exhibits high specific activity and retained complex stability, while other variants were either significantly less active (< 30% of wild type) and/ or incomplete regarding their subunit composition. These findings were basically confirmed with the enzymes from *C. acetobutylicum* and the plasmid *pASG.3´_bcdABC* was used for recombinant production of Butyryl-CoA dehydrogenase from *C. difficile* and *C. acetobutylicum* in *E. coli*. The purification was performed under both aerobic and anaerobic conditions. As shown by SDS-PAGE (Fig.4.23 & 4.24), three bands between 28 and 43 kDa corresponding to purified butyryl-CoA dehydrogenase/ETF complex were observed and it is almost equal to the mass calculated from amino acid sequence. Staining intensities suggest a stoichiometry ($\alpha\beta\gamma$) of 1:1:1 and 1:2:1 for butyryl-CoA dehydrogenase/ ETF complex from *C. difficile* and *C. acetobutylicum*

produced in *E. coli* after purification under air, respectively. Under anaerobic conditions, the same stoichiometry ratio for *C. difficile* BCD/ETF complex was observed. In contrast, the *C. acetobutylicum* complex was un-stable and only one band related to the *C*-terminal tagged ETFA subunit was obtained. The results obtained from both aerobic and the anaerobic conditions suggested that the BCD/ETF complex from *C. acetobutylicum* was unstable.

The enzyme activity was measured in both the oxidation and reduction direction to facilitate understanding of the reaction mechanism.

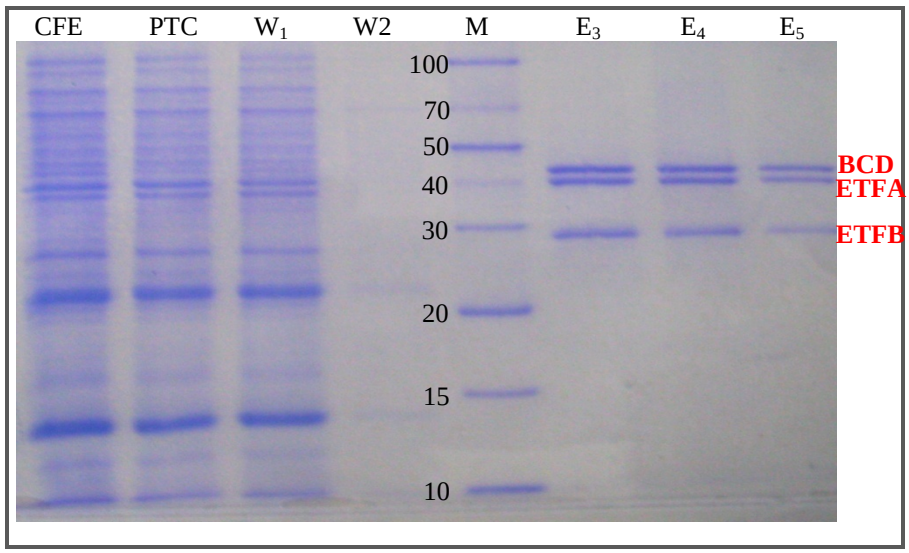


Fig.4.23: SDS-PAGE 15% showing purity and molecular mass for butyryl-CoA dehydrogenase/ ETF complex from *C. difficile* produced in *E. coli* after purification under air.

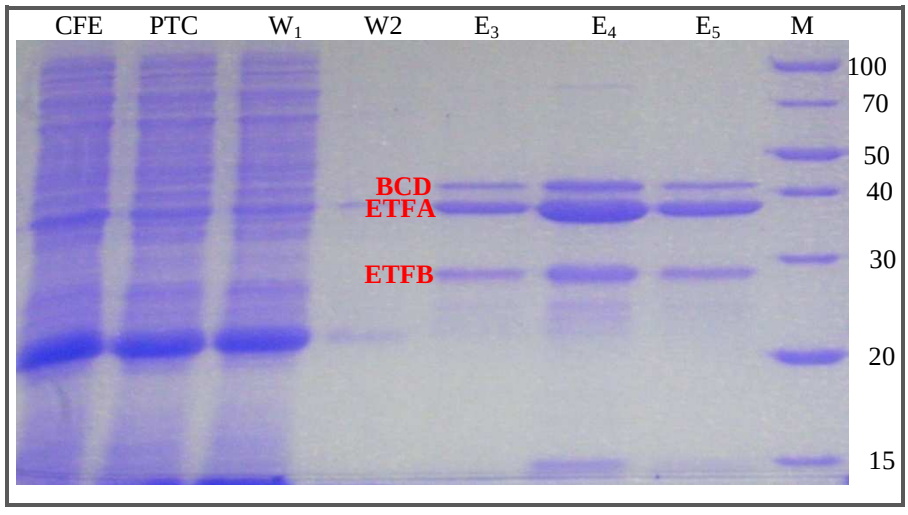
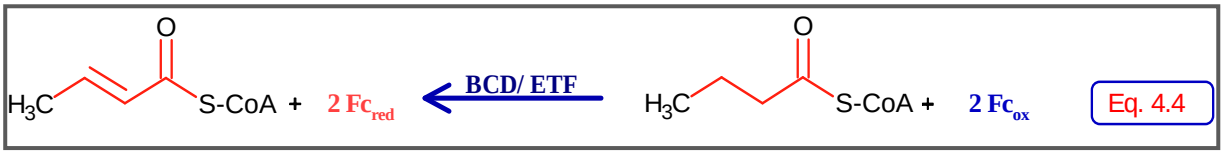


Fig.4.24: SDS-PAGE 12% showing purity and molecular mass for butyryl-CoA dehydrogenase/ ETF complex from *C. acetobutylicum* produced in *E. coli* and purificed under aerobic conditions.

4.2.4.1. Ferricenium hexafluorophosphate assay

Butyryl-CoA dehydrogenase/ ETF complex from *C. difficile* is active under aerobic conditions and it is not sensitive to oxygen. In contrast, the one from *C. acetobutylicum* is slightly oxygen sensitive **[56-59]**. The activity was measured under aerobic conditions using ferricenium hexafluorophosphate ($Fc^+PF_6^-$) as artificial electron acceptor (Eq. 4.4), in an assay which it is well established for measuring activity of acyl-CoA dehydrogenases in the oxidative direction. Enzyme activity was calculated based on monitoring the reduction of ferricenium at 300 nm using a molar extinction coefficient of 8.6 (2*4.3) $mM^{-1}cm^{-1}$ **[101]**. The assay contained Kpp (pH 7.5, 50 mM), FAD (5.0 µM), ferricenium (100 µM) and butyryl-CoA (50 µM). The specific activities for the BCD/ ETF complexes obtained from *C. difficile* and *C. acetobutylicum* in CFE were 0.41 and 0.72 U/mg, respectively, for the complexes with *C*-terminal tagged ETFA-subunits. After purification, specific activities raised to 10.2 and 8.4 U/mg achieving 25 and 12-time fold purification, respectively.



Kinetic Constants: The apparent K_M and V_{max} values for butyryl-CoA were calculated from non-linear regression of the Michaels-Menten equation (Fig. 4.25). Butyryl-CoA dehydrogenase from *C. difficile* has an apparent K_M of 10 µM for butyryl-CoA at an apparent V_{max} of 15.3 U/mg. It was a result of testing varied butyryl-CoA concentrations (2.5- 35 µM) in the presence of fixed concentration of ferricenium (100 µM) and 180 ng pure enzyme complex. The turnover number (k_{cat}) for butyryl-CoA dehydrogenase from *C. difficile* calculated from a sub-complex molecular mass of 110.3 kDa was 28 s^{-1} per active site and the k_{cat}/ K_M was 2.8 $\mu M^{-1}s^{-1}$.

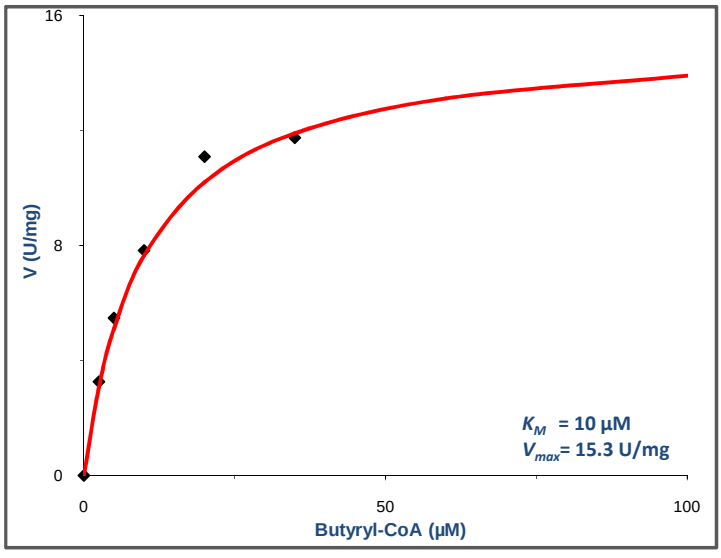


Fig.4.25: Michaels-Menten plots of BCD/ ETF complex from *C. difficile* produced in *E. coli*.

4.2.4.2. NADH oxidation (aerobic assay)

The rate of NADH oxidation of the *C. difficile* BCD/ ETF complex under oxic conditions was determined based on monitoring the decrease at 340 nm using a molar extinction coefficient of 6.22 $mM^{-1}cm^{-1}$ **[57]** in an assay containing Tris-HCl (pH 7.5, 50 mM), $MgCl_2$ (10 mM), NaCl (50 mM) and NADH (150 µM). The enzyme had a diaphorase activity (Eq. 4.5) of 0.22 U/mg. The effect of different substrates like FAD, oxidized Ferredoxin (Fd_{ox}) and Crotonyl-CoA on the oxidation rate of NADH was determined and the data were plotted in figure (4.26).

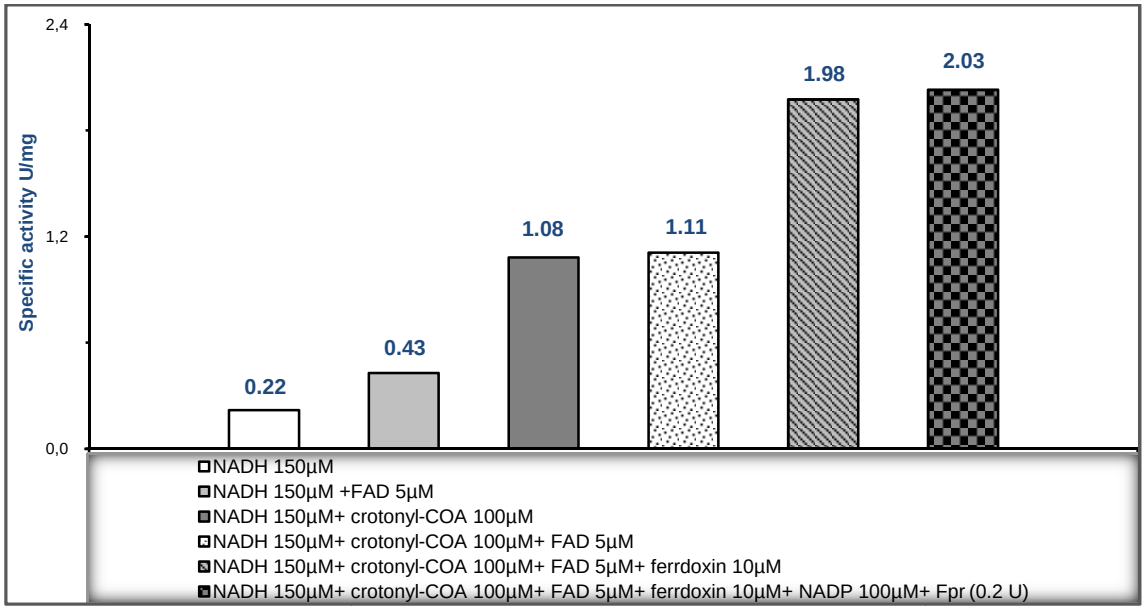


Fig.4.26: Effect of different substrates on the NADH oxidation by BCD/ ETF complex from *C. difficile* produced in *E. coli*. FPR is Ferredoxin: NADP oxidoreductase (EC 1.18.1.2) from *E. coli*.

4.2.4.3. NADH / Crotonyl-CoA reduction (anaerobic assay)

With an attempt to measure the enzyme in the physiological direction with crotonyl-CoA and NADH under anoxic conditions, different assay conditions were tested and the data were plotted in figure (4.27) and table (4.4). The assay contained Tris-HCl (pH=7.5, 50 mM), $MgCl_2$ (10 mM), NaCl (50 mM), NADH (150 µM), Crotonyl-CoA (50 µM) and 200 ng purified BCD/ ETF complex. The highest activity was monitored when FAD, Fd_{ox} as well as the regeneration system for oxidized ferredoxin was used (Eq. 4.6). It should be noted that all of the following measurements were performed only for the *C. difficile* BCD/ ETF complex because the complex from *C. acetobutylicum* was incomplete and nopurification for the three subunits (αβγ) could be achieved under anoxic conditions which made the enzyme inactive.

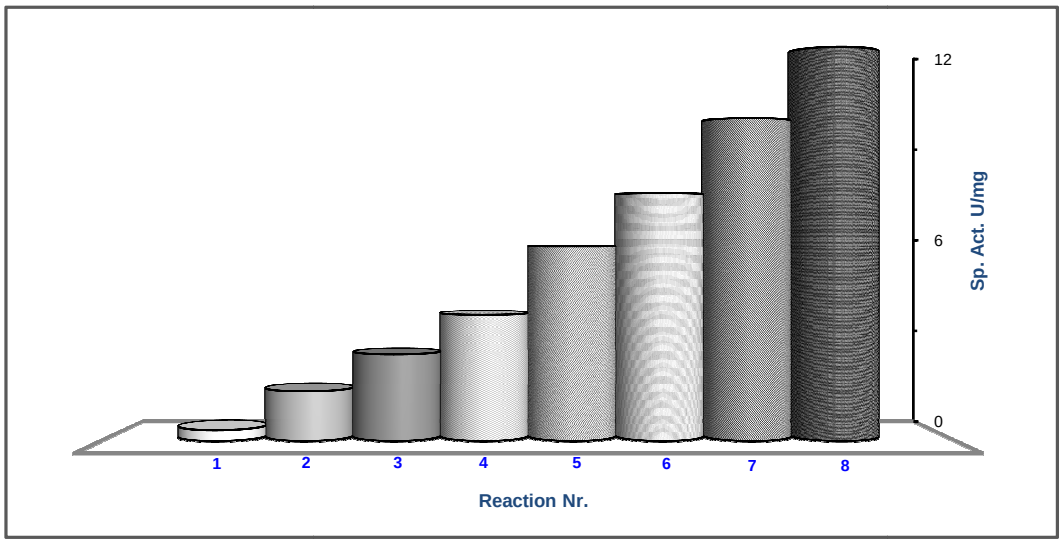
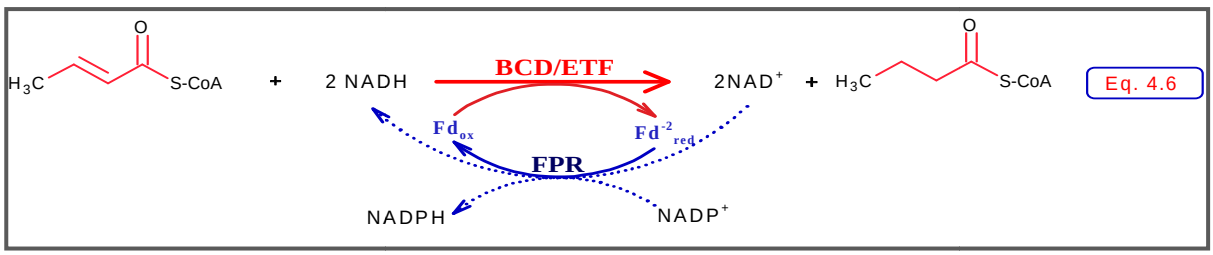


Fig.4.27: Establishment the optimum measurement for BCD/ ETF complex from *C. difficile*

Table.4.4: Establishment the optimum measurement for BCD/ ETF complex

Reaction Nr.	FAD	Fd	FPR	NADP	Sp. Act. (U/mg)	Enrichment
1	**0 µM**	-----	-----	-----	0.33	1
2	**5 µM**	-----	-----	-----	1.48	5
3	**10 µM**	-----	-----	-----	2.58	8
4	10 µM	1 µM	-----	-----	3.76	12
5	10 µM	1 µM	**0.3 U**	-----	5.8	18
6	10 µM	**1 µM**	0.3 U	**100 µM**	7.39	23
7	10 µM	**2 µM**	0.3 U	100 µM	9.63	30
8	10 µM	**5 µM**	0.3 U	100 µM	11.75	37

The varied factors are in bold blue font.



Kinetic Constants of BCD/ ETF complex: The apparent K_M, V_{max}, k_{cat} and k_{cat}/K_M values for NADH, Ferredoxin and Crotonyl-CoA in the physiological direction (Eq.4.6) were calculated from non-linear regression of Michaels-Menten equation (Fig. 4.28, 29 & 30) and the data was presented in table (4.5).

Table.4.5: Kinetic constants of BCD/ ETF complex

Substrate	K_M (µM)	V_{max} (U/mg)	Sub-complex mass (kD)	k_{cat} (s^{-1})	k_{cat}/K_M ($\mu M^{-1}s^{-1}$)
NADH	145	19.4	110.3	36	0.25
Ferredoxin	2.0	20.2	110.3	37	18.5
Crotonyl-CoA	2.5	10.2	110.3	19	7.6

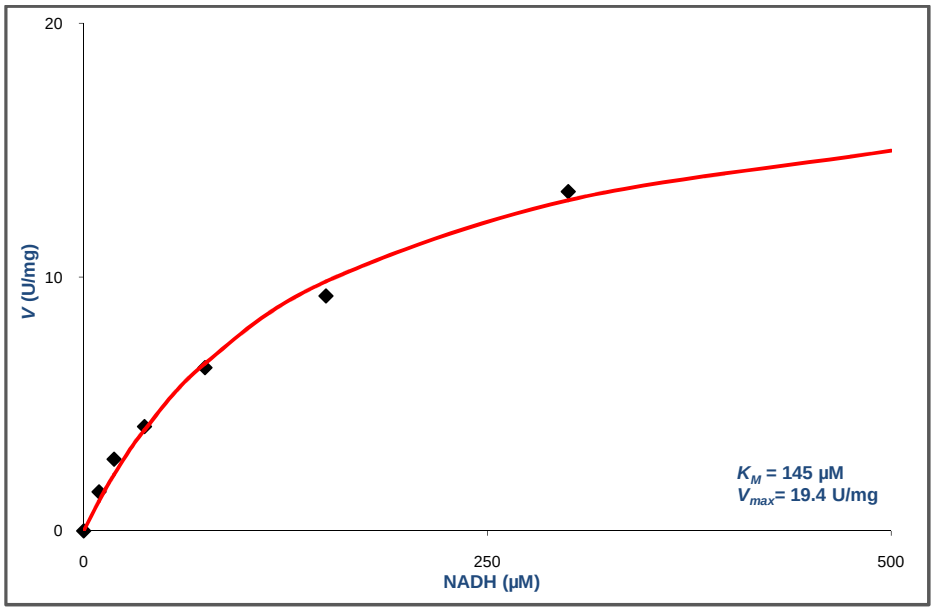


Fig.4.28: Michaels-Menten plots for BCD/ ETF complex from *C. difficile* produced in *E. coli*. The assay contained Tris-HCl (pH=7.5, 50 mM), $MgCl_2$ (10 mM), NaCl (50 mM), FAD (10 µM), NADP (100 µM), FPR (0.3U), 120 ng pure BCD/ETF, Fd (10 µM) and Crotonyl-CoA (50 µM).

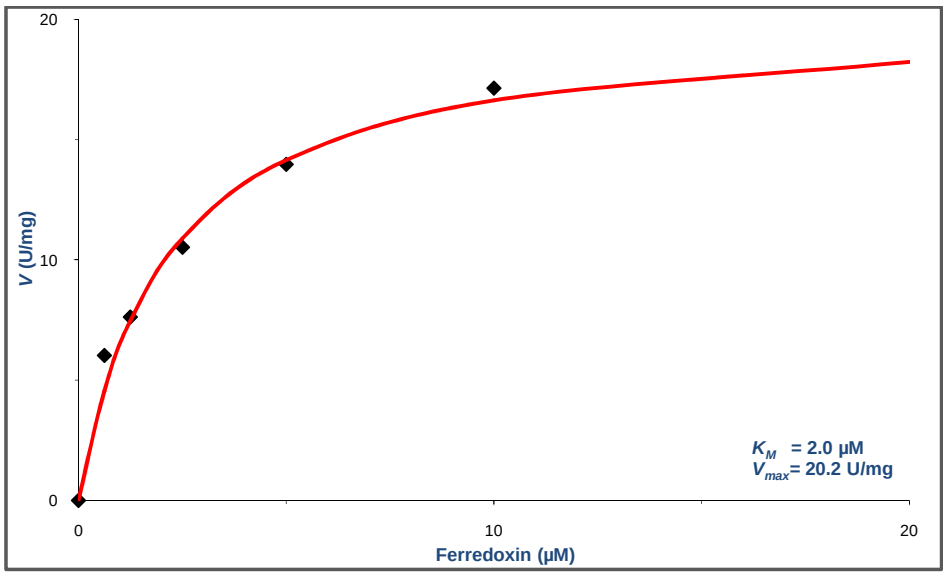


Fig.4.29: Michaels-Menten plots for BCD/ ETF complex from *C. difficile* produced in *E. coli*. The assay contained Tris-HCl (pH=7.5, 50 mM), $MgCl_2$ (10 mM), NaCl (50 mM), FAD (10 µM), NADP (100 µM), FPR (0.3U), 120 ng pure BCD/ETF, NADH (200 µM) and Crotonyl-CoA (50 µM).

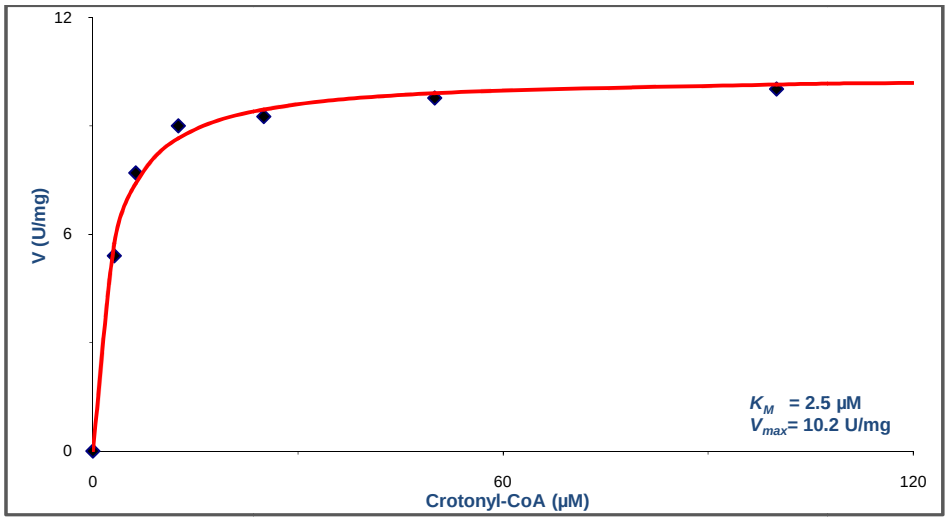


Fig.4.30: Michaels-Menten plots for BCD/ ETF complex from *C. difficile* produced in *E. coli*. The assay contained Tris-HCl (pH=7.5, 50 mM), $MgCl_2$ (10 mM), NaCl (50 mM), FAD (10 µM), NADP (100 µM), FPR (0.3U), 120 ng pure BCD/ETF, Fd (10 µM) and NADH (200 µM).

The re-generation system employed in the current study was Ferredoxin: NADP reductase (FPR, EC 1.18.1.2) which couples the reduction of NAD(P) with reduced ferredoxin (Fd_{red}) to form NAD(P)H and oxidized ferredoxin (Fd_{ox}). The *fpr* gene was cloned from *E. coli* (accession Nr. b3924) and expressed under *tet*-promoter control. The protein was purified and characterized in the current work (Fig.4.31). The molecular mass for the *C*–terminally tagged protein was 29 kDa. The specific activity for the pure enzyme was 38 U/mg in the assay containing Tris-HCl (pH=7.5, 50 mM), $MgCl_2$ (10 mM), NaCl (50 mM), NADPH (200 µM) and Ferricyanide (100 µM). The enzyme had 23% activity with NADH compared to NADPH and this activity decreased in the presence of FAD. The V_{max} was about 56 U/mg and K_M of 18 and 11 µM for NADPH and Ferricyanide were obtained, respectively (Fig.4.32).

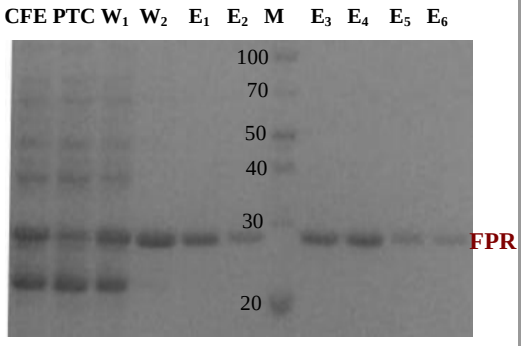


Fig.4.31: SDS-PAGE 15% showing purity and molecular mass for FPR from *E. coli*.

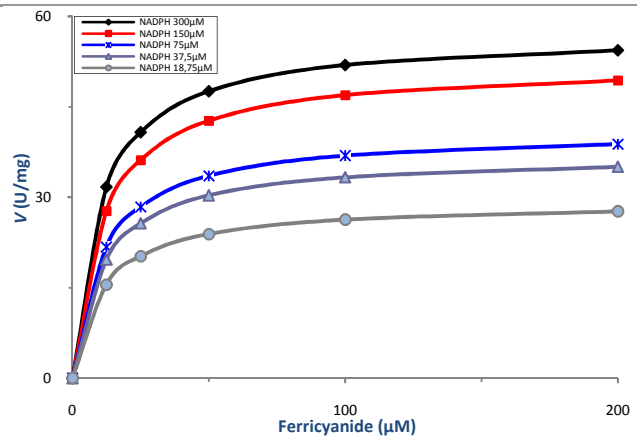


Fig.4.32: Michaels-Menten plots for FPR.

4.2.5. Butyrate kinase

Butyrate kinase (EC 2.7.2.7), which catalyzes the reversible formation of butyrate from butyryl phosphate with concomitant phosphorylation of ADP, has been demonstrated in different clostridia **[27, 43, 45-47]**. The purified *C*-terminally tagged enzyme showed a single subunit band corresponding to a molecular mass of 42.6 kDa (Fig.4.33), which is in agreement with its calculated molecular mass (40 kDa). The native molecular mass was determined by Native-PAGE and it was about 89± 3 kDa (Fig.4.35). This suggests that the native enzyme is apparently a homodimer (2*42.6).

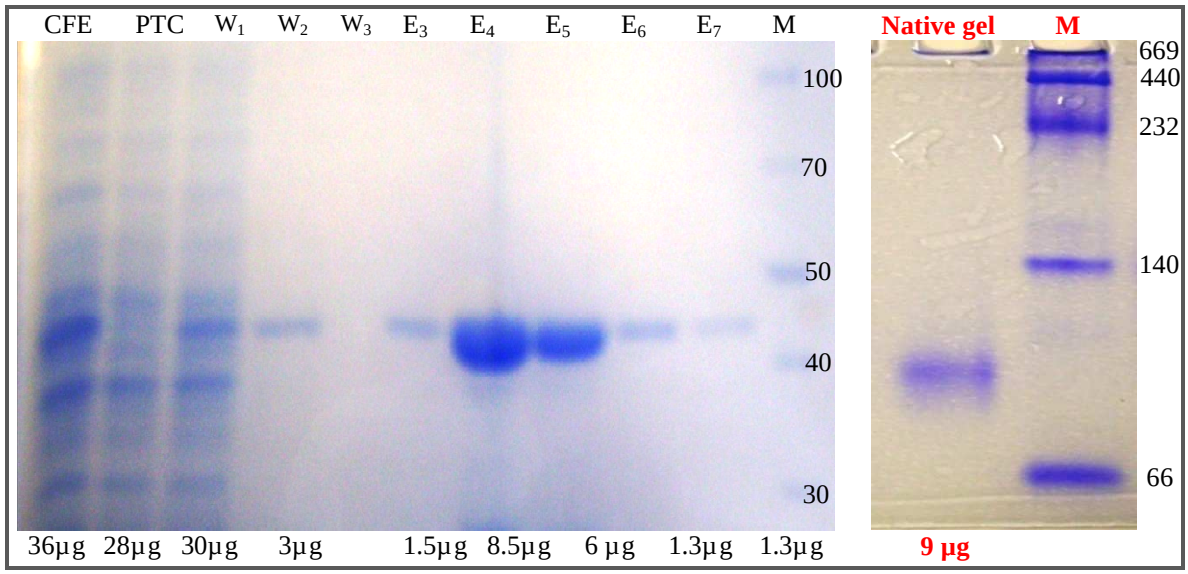
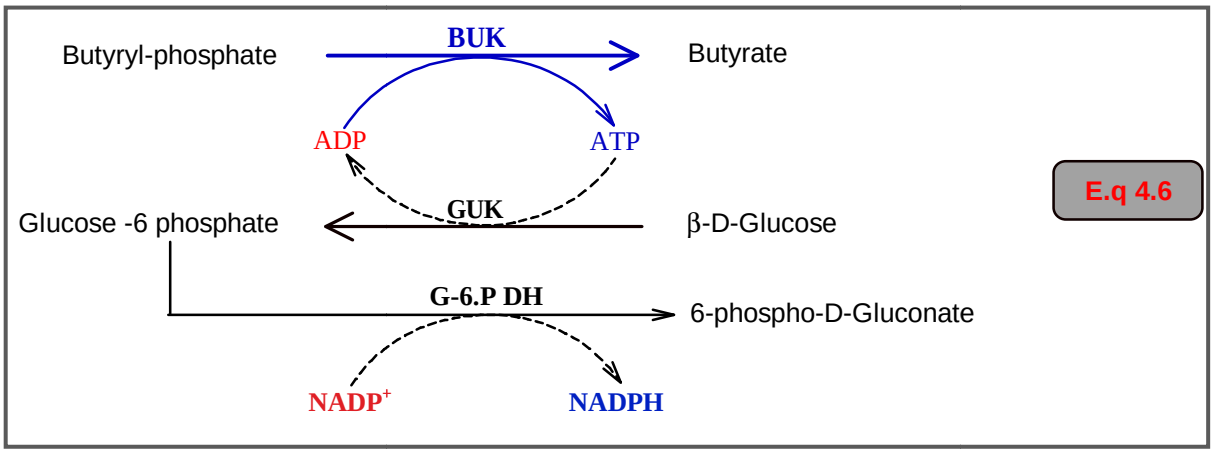


Fig.4.33: Purity and molecular mass of *C*-terminally tagged BUK.

The left gel is 12% SDS-PAGE Stained with Coomassie.

The right gel is 7% Native-PAGE showing the native molecular mass of BUK.



In the recent investigation, a novel method to measure the activity of BUK in the physiological direction (the butyrate forming direction) was developed (Eq. 4.6). The assay was coupled with glucokinase (EC 2.7.1.2) and glucose 6-phosphate dehydrogenase (EC 1.1.1.49). The increase in absorbance at 340 nm due to the reduction of NADP by the latter

enzyme was monitored using a molar extinction coefficient of 6.22 $mM^{-1}cm^{-1}$ **[102]**. The assay contained Kpp (pH 7.5, 50 mM), $MgCl_2$ (5 mM), glucose (4 mM), (1 mM) DTT, $NADP^+$ (200 µM), ADP (100 µM), butyryl-phosphate (1 mM), glucokinase and glucose 6-phosphate dehydrogenase (one unit each). The specific activity in cell free extract was 21 U/ mg and after purification, it had sharply increased to achieve 322 U/mg with about 15-time fold.

Kinetic Constants. Kinetic properties of BUK were analyzed in the butyrate-forming direction by varied butyryl-phosphate concentrations (0.09- 5.75 mM) in the presence of different fixed concentrations of ADP (25- 250 µM) and 43 ng pure enzyme (Fig.4. 34). In order to calculate K_M and V_{max} values graphically, Lineweaver-Burk plots were employed (Fig.4.35). The straight lines were intersecting in *x*-axis showing an apparent K_M value of 470 µM for butyryl-phosphate (Fig.4.35). Moreover, a secondary graph (Fig.4.36) was plotted with the reciprocal of the ADP concentration against the ordinate intercepts, which were obtained from figure (4.35). The enzyme had an apparent K_M for ADP of about 70 µM with an apparent V_{max} of 345 U/mg. The calculated turnover number (k_{cat}) was 245 s^{-1} and k_{cat}/ K_M was 0.52 $\mu M^{-1}s^{-1}$. The Lineweaver-Burk plots (Fig.4.35) indicate that the butyrate kinase reaction follows a formation of a ternary complex of enzyme-butyrylphosphate-ADP by random *bi-bi* pattern.

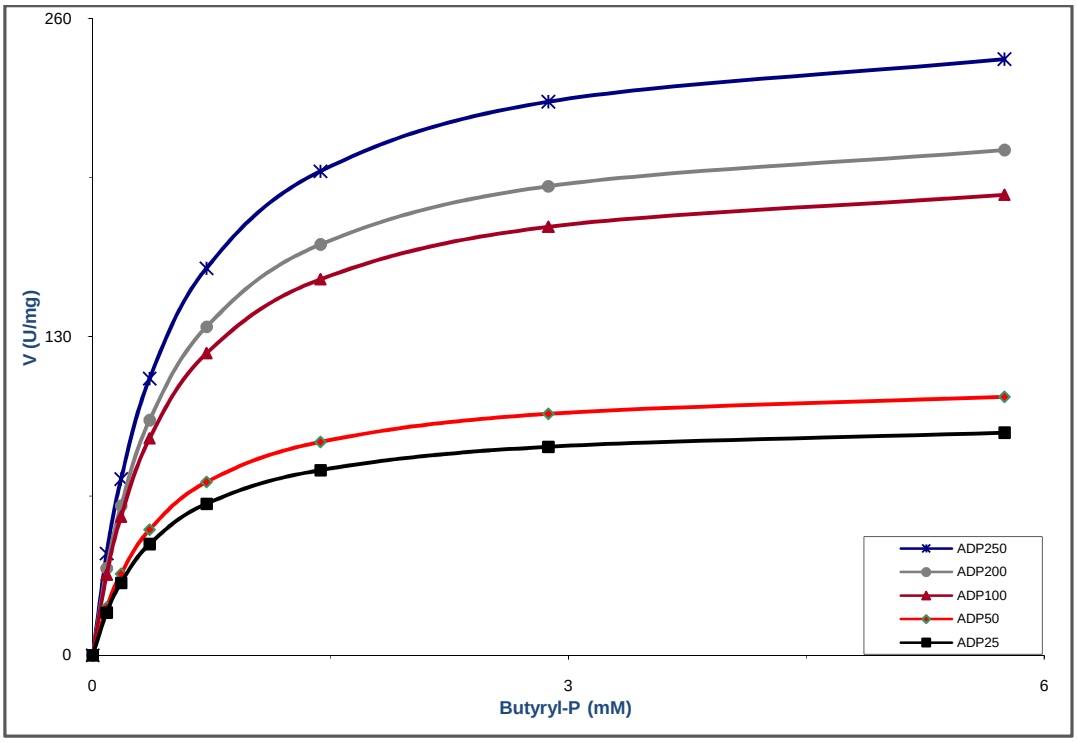


Fig.4.34: Michaels-Menten kinetics of BUK.

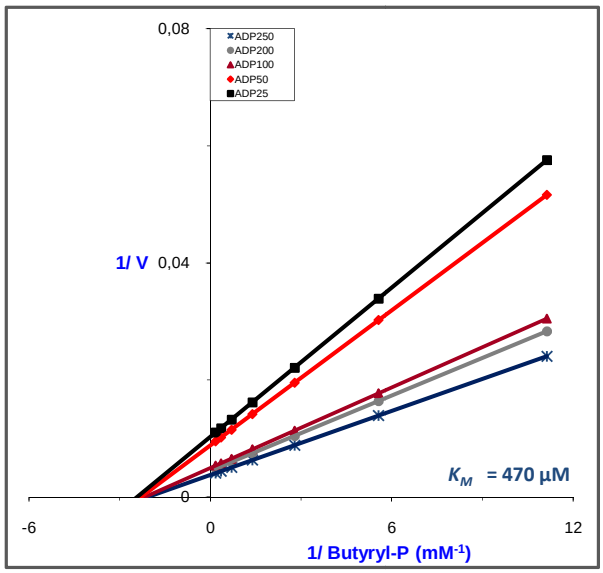


Fig.4.35: Lineweaver-Burk plots of BUK at varied butyryl-phosphate. The legend shows the ADP concentrations.

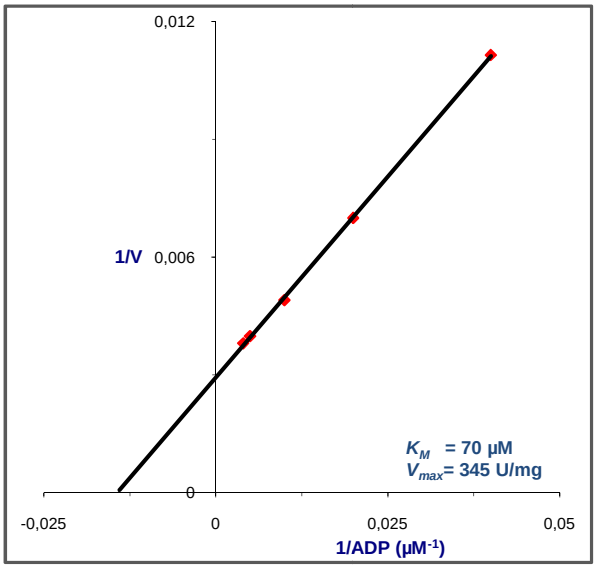


Fig.4.36: Secondary plots of ADP against ordinate intercepts which were derived from figure (4.35).

4.2.6. Phosphate butyryltransferase (PBT)

Phosphate butyryltransferase (EC 2.3.1.19), an enzyme capable of catalyzing the reversible transfer of the butyryl group from butyryl-phosphate to CoA-SH, forming butyryl-CoA and an inorganic phosphate, has been found in all clostridia. The purified *N*-terminally tagged PBT showed a single subunit band corresponding to a molecular mass of 34.2 kDa (Fig.4.37) and is in agreement with the predicted molecular mass (33.7 kDa).

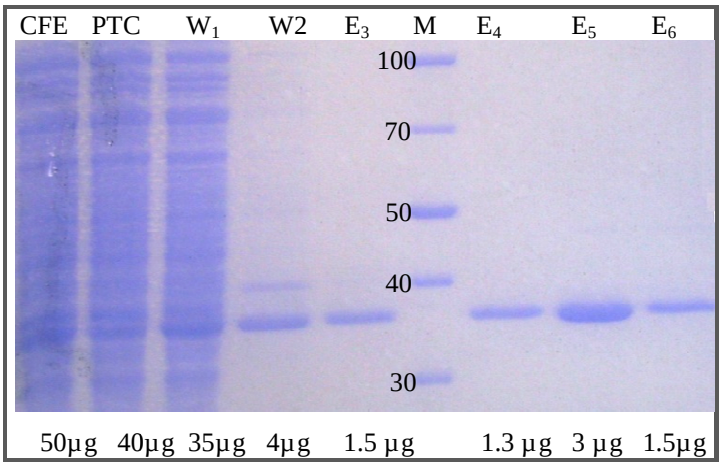
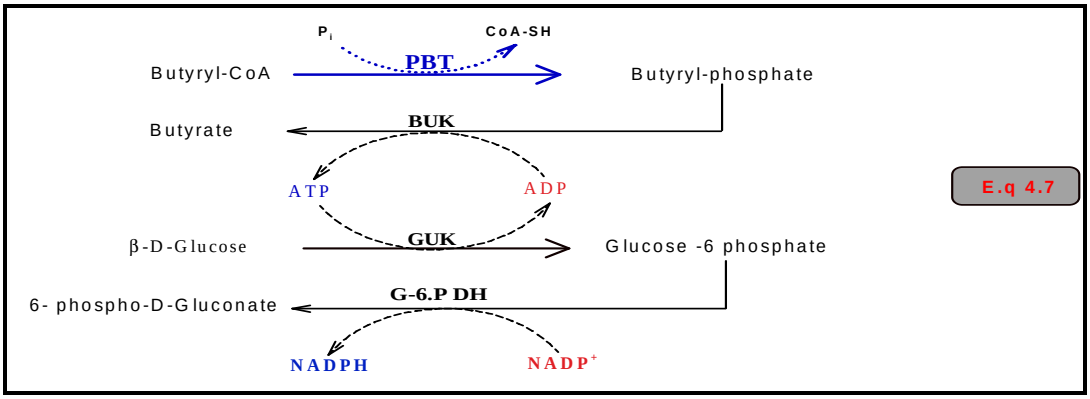


Fig.4.37: Purity and molecular mass of *N*-terminally tagged PBT on 12% SDS-PAGE.

The above mentioned novel method for butyrate kinase was adapted to measure the activity of PBT in the physiological direction (the butyryl-phosphate forming direction) (Eq. 4.7). The assay was coupled with recombinant butyrate kinase, glucokinase and glucose

6-phosphate dehydrogenase. This method was more effective than the commonly used DTNB assay, because the latter is strongly affected by DTT, which was used as an anti-oxidant to protect enzyme activity. The assay contained Kpp (pH 7.5, 50 mM), $MgCl_2$ (5 mM), glucose (4 mM), $NADP^+$ (200 µM), ADP (100 µM), butyryl-CoA (50 µM), DTT (1 mM) and one unit of each glucokinase, glucose 6-phosphate dehydrogenase and butyrate kinase. The specific activity was 8 U/mg in the cell free extract and 180 U/mg for the pure protein (22 fold). The enzyme activity with alternative substrates was 5% with acetyl-CoA, 26% with acetoacetyl-CoA, and 33% with propionyl-CoA.



Kinetic Constants. Apparent K_M and V_{max} values for PBT from *C. difficile* produced in *E. coli* were calculated mathematically with non-linear regression of Michaels- Menten equation and graphically from Lineweaver-Burk plot (Fig.4.38) in the previously described buffer, different butyryl-CoA concentrations (1.5- 200) and 50 ng pure enzyme. The enzyme had apparent K_M of 72 µM for butyryl-CoA with apparent V_{max} 257 U/mg. At optimal substrate concentration the turnover number (k_{cat}) was 151 s^{-1} and the k_{cat}/K_M was 2.1 $\mu M^{-1}s^{-1}$.

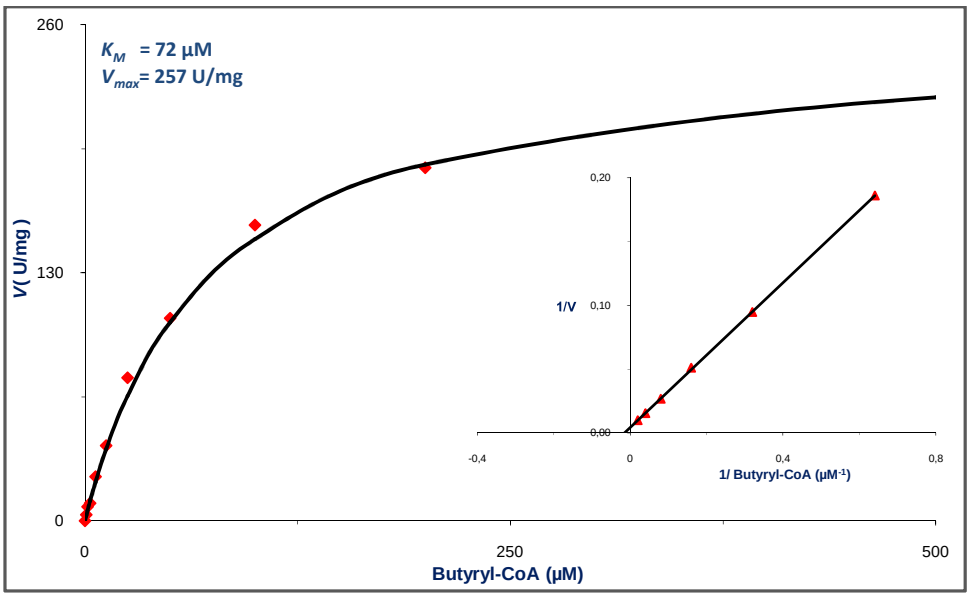


Fig.4.38: Michaels-Menten and Lineweaver-Burk plots of phosphate butyryltransferase activity.

4.3. Establishment of the butyrate synthetic pathway in *Escherichia coli*

4.3.1. Construction of co-expression plasmids

Since it has been previously shown that the individual enzymes required for butyrate production were produced in *E. coli* in an active state, two different butyrate metabolic pathway modules, *pASG.wt_BUT*$_D$ and *pASG.wt_BUT*$_{AC}$, are constructed (the construction scheme is given in figure 4.39). The underlying fusion process was previously exemplified with butyryl-CoA dehydrogenase/ ETF complex (Fig.4.8 & 4.9).

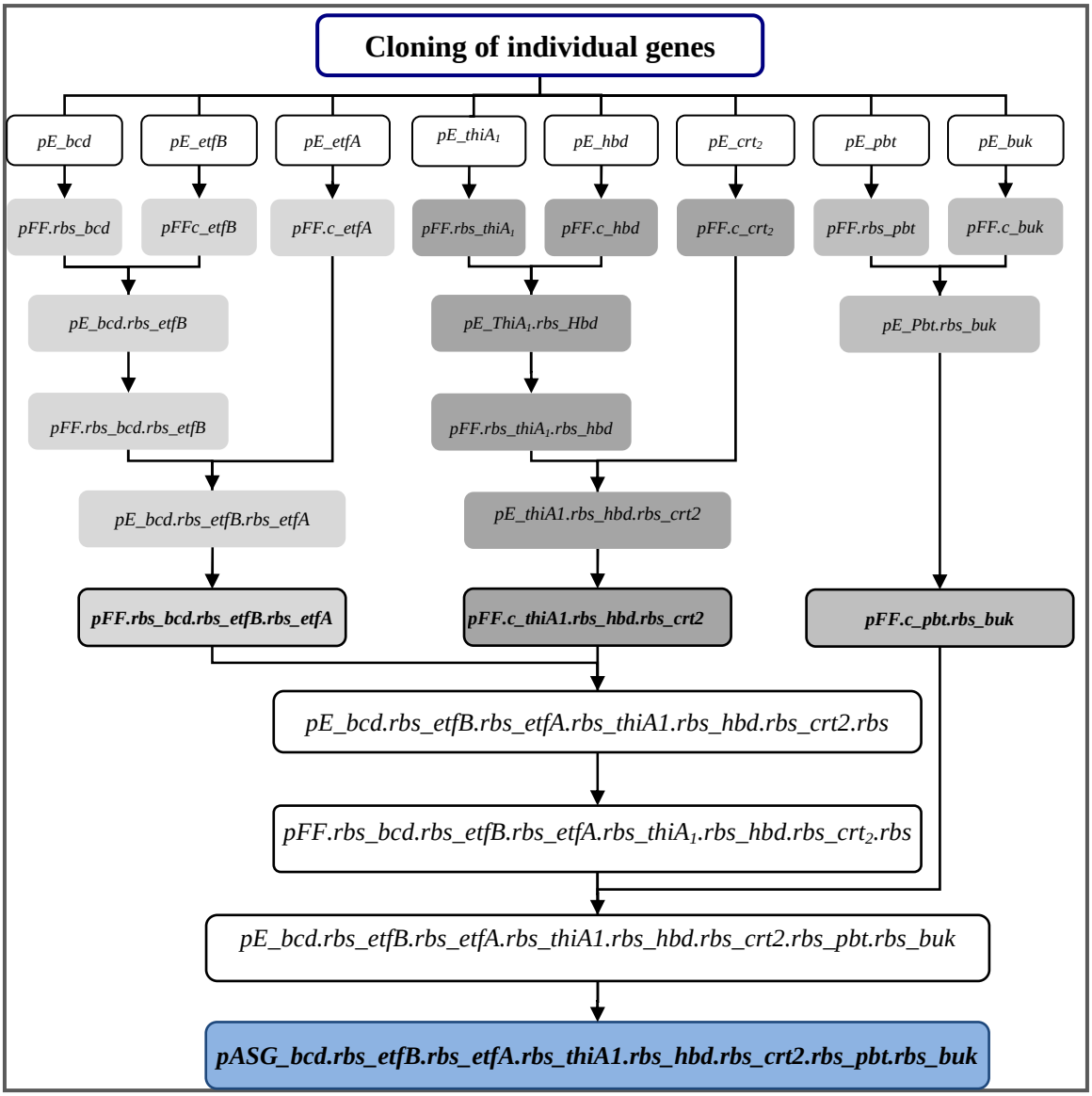


Fig.4.39: Construction scheme of butyrate metabolic pathway module.

Please note that both expression plasmids contained most butyrate fermentation genes (*thiA*$_1$, *hbd*, *crt*$_2$, *pbt* and *buk*) from *Clostridium difficile*. Only the butyryl-CoA dehydrogenase/ETF complex genes originated from either *C. difficile* (*pASG.wt_BUT*$_D$) or *C. acetobutylicum* (*pASG.wt_BUT*$_{AC}$) (Fig.4.40)

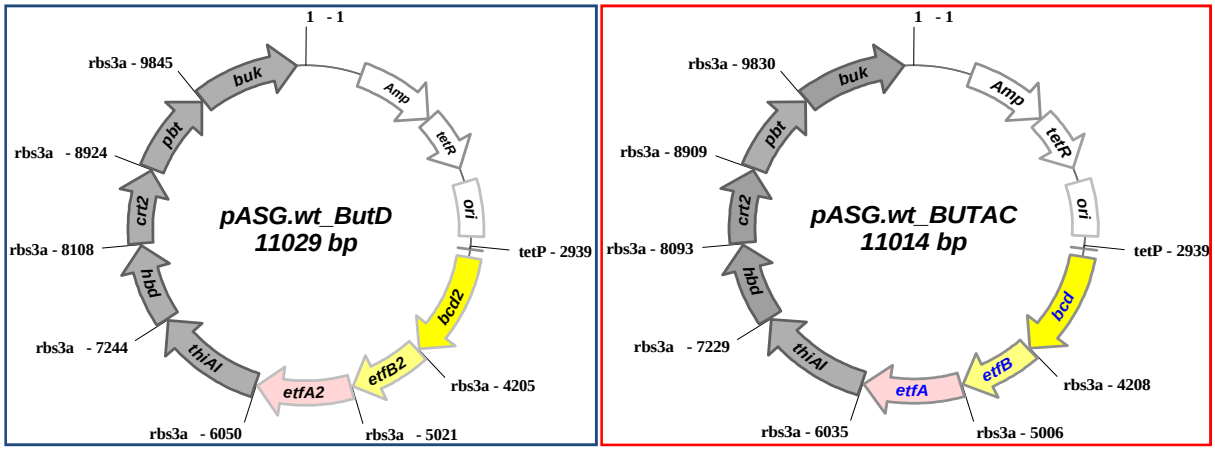


Fig.4.40: The constructed butyrate biosynthetic expression plasmids used in the current work.

The plasmids were then transformed into *E. coli* Rosetta and the expression of the foreign genes was accomplished by induction of the *tetR*-regulated artificial operon by anhydrotetracycline (AHT, 200 ng/ mL).

As summarized in tables (4.6 and 4.7), the recombinant proteins were readily detected in cell-free extracts of the recombinant strains (further referred to as BUT_D and BUT_{AC}), while their activities were either totally absent or very low in control cells. In fact, it turned out that the expression levels of individual enzymes was strongly reduced as compared with the single enzyme expression for purification of particular targets, but the enzyme production still remained clearly visible on Coomassie-stained SDS-PAGE (Fig.4.41.b). Thus, it can be stated that also under the conditions of co-expression some enzymes required for the transfer of acetyl-CoA to butyrate were produced in similar amounts and clearly detectable via the enzyme activities.

Table 4.6: Enzymes specific activities in *E. coli* Rosetta, BUT_D and BUT_{AC} strains

strains	Specific activity (U/mg)					
	$THIA_1$	BHBD	CRT_2	BCD	PBT	BUK
Rosetta	0.4	>>0.01	0.2	*ND*	*ND*	0.1
BUT_D	13.8	0.38	64.3	0.043	1.95	13.2
BUT_{AC}	18.7	0.65	89.1	0.098	2.09	25.2

Table 4.7: The expression levels of individual genes in recombinant *E. coli* strains

strains	$THIA_1$	BHBD	CRT_2	BCD	PBT	BUK
BUT_D	12	3	1	4	3	19
BUT_{AC}	11	3	1	*Nd*	3	37

ND, not detected; Nd, not determined

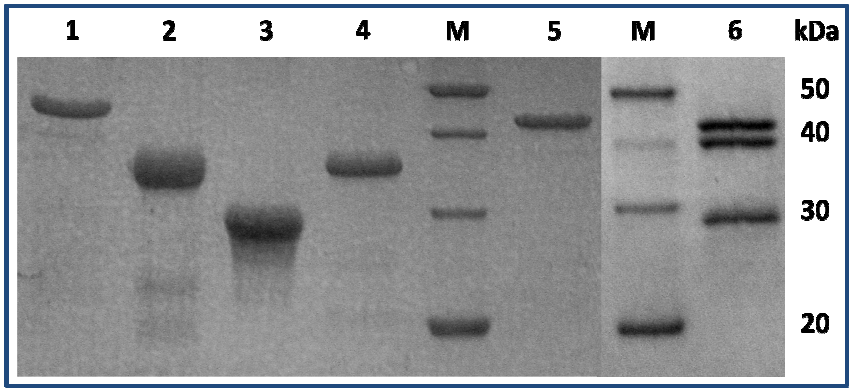


Fig.4.41.a: Coomassie-stained gels with individually purified recombinant enzymes. 5-10 µg of C-terminally StrepII-tagged $THIA_1$ (45 kDa, lane 1), BHBD (34 kDa, lane 2), CRT_2 (30 kDa, line 3), BUK (43 kDa, lane 5) and N-terminally StrepII-tagged PBT (35 kDa, lane 4) were separated on 12% SDS-PAGE. The BCD/ETF complex (15 µg) composed of the reductase (42 kDa) and the ETF-subunits A (39 kDa) and B (29 kDa) with a C-terminal StreppII-tag attached to ETFA was separated on 15% SDS-PAGE. The sizes of relevant proteins of PageRuler™ (Fermentas) molecular mass standards (M) are indicated for comparison.

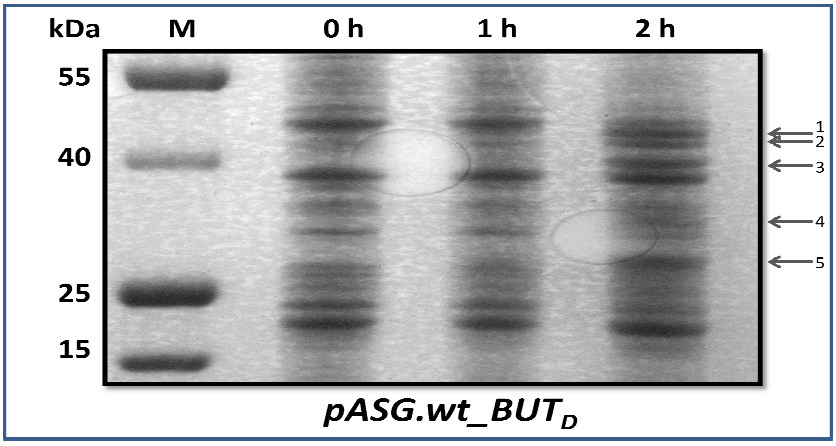


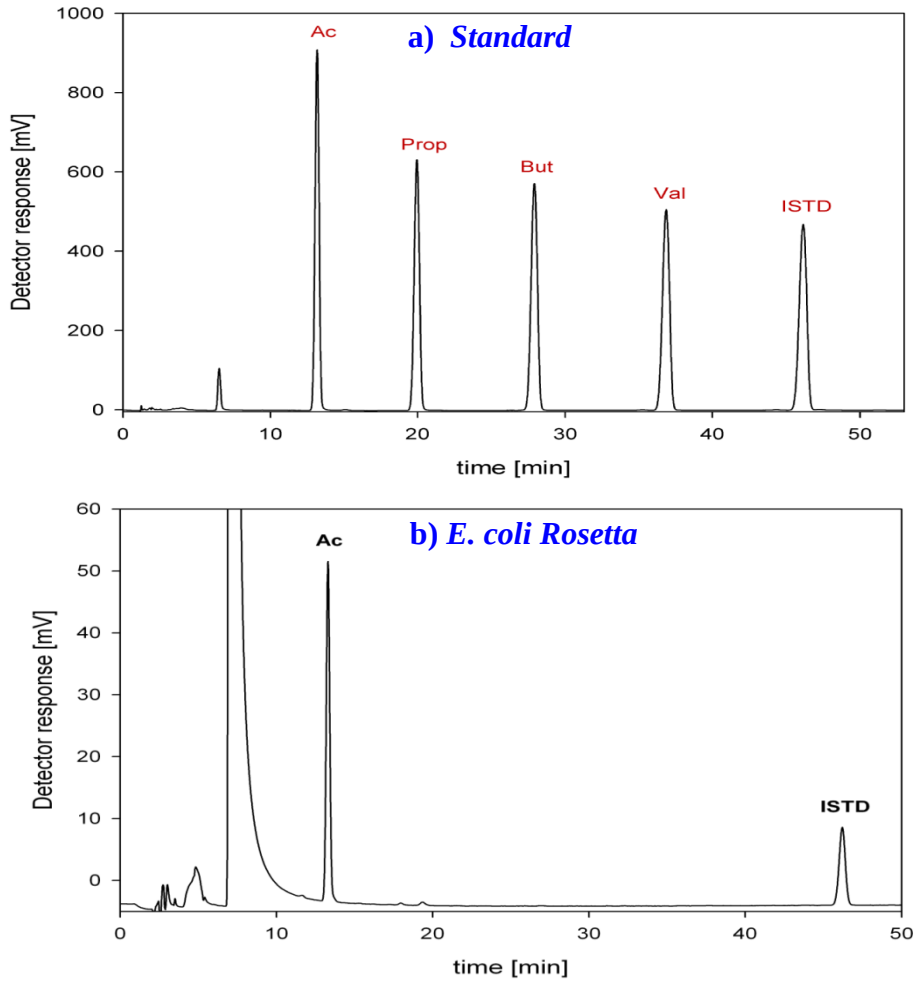
Fig.4.41.b: Coomassie-stained gels show the expression of butyrate pathway biosynthetic module (*pASG.wt-BUT$_D$*). The induction times (per hours) indicate above each lane. The right arrows show the new bands corresponding to 1, $THIA_1$ plus BUK; 2, BCD; 3, ETFA; 4, BHBD; 5, CRT_2 plus ETFB. The sizes of relevant molecular mass standards (M) are indicated for comparison.

4.3.2. Production of butyrate by the recombinant strains

The recombinant production of the butyrate forming enzymes in *E. coli* encouraged us to investigate the formation of its production by the recombinant strains. As previously known, *E. coli* does neither produce butyric acid as a fermentation end product nor is it capable of utilizing butyrate as carbon/energy source **[90, 108]**. The introduction of the foreign genes was predicted to induce both capabilities in the recombinant strains. *E. coli* Rosetta and the recombinant strains were checked for butyrate production upon growth on M_9 liquid medium with 20 mM glucose as a sole carbon source under aerobic conditions

(Fig.4.42.b, c & d). From these cultures, samples were taken at various time-points and analyzed in order to collect data sets including growth, substrate consumption and product formation.

The biomass (dry mass) of *E. coli* was calculated from the optical density of the culture with a correlation factor of 0.35 mg dry mass per unit OD_{578nm}. The glucose concentration in the medium was followed by a coupled optic enzymatic test using glucokinase from *C. difficile* and glucose-6-phosphate dehydrogenase from *Bacteroides fragilis* (both recombinantly produced in *E. coli*). The fatty acid analysis was performed using a HPLC-method recently developed in our laboratory **[109]**. Therefore, fatty acids were extracted from the medium with diethyl ether and converted to the corresponding acid chlorides with oxalylchloride. The acid chlorides were then esterified with *p*-nitrophenol and the nitrophenyl esters were separated by RP-HPLC using a Synergi Polar-RP (18Å, 4µm, 250*4.6 mm) column as stationary phase. As shown in figure (4.42.a), fatty acids from C_2 to C_5 were readily separated by the method employed and allowed simultaneous quantification of acetic, propionic, butyric and valeric acid using an internal standardization calibration with cyclohexyl carboxylic acid as internal standard.



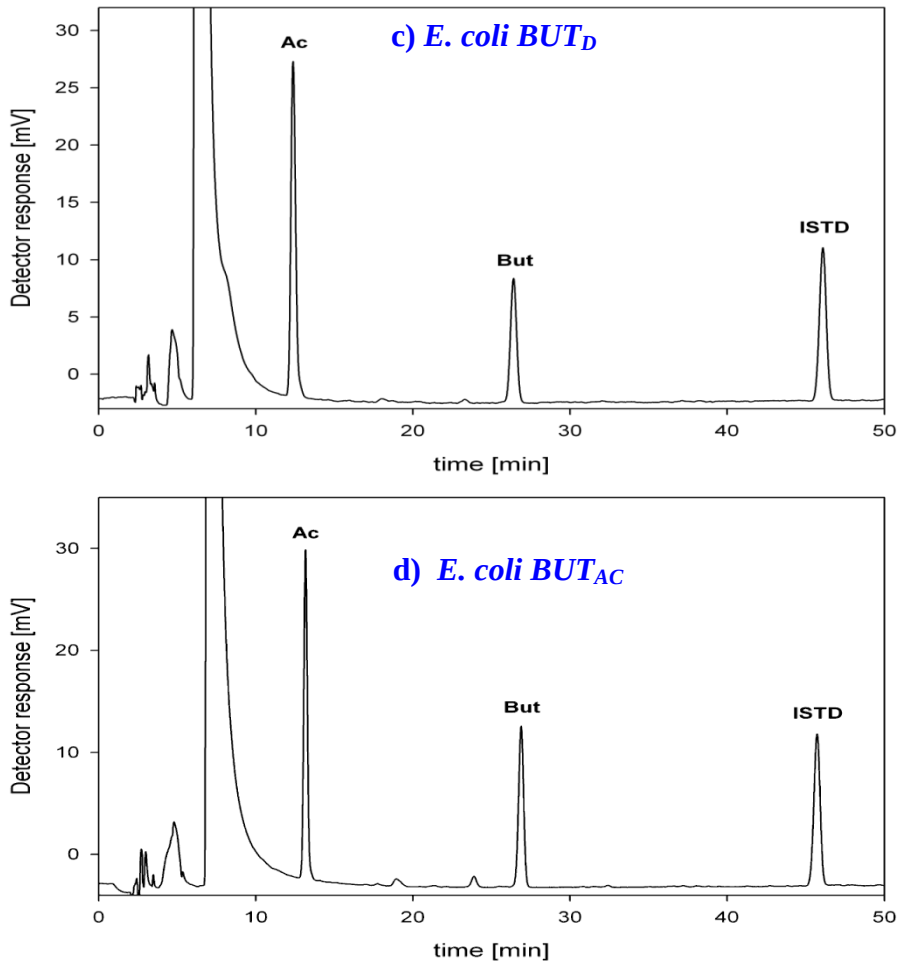


Fig.4.42: Chromatograms show the analysis of short-chain fatty acids: a) Standard of 20 mM of Ac, acetate; Pro, propionate; But, butyrate; Val, valerate; ISTD, cyclohexyl carboxylic acid was used as internal standard. b) *E. coli* Rosetta obtained as control; c) *E. coli* BUT_D; d) *E. coli* BUT_{AC}.

As expected, the host strain as well as the recombinant ones readily consumed glucose and grew reasonably. The fatty acid analysis revealed a growth-associated increase of the acetate concentration during exponential growth, which reached a maximum of 16.6 mM in the host or 9.3 and 6.6 mM in the BUT_D- or BUT_{AC}-strains, respectively. As expected, only the recombinant strains were capable to form butyrate and maximum values of 3.1 and 3.7 mM were obtained, respectively.

As shown in figure (4.43.a), the host strain consumed the offered glucose rapidly and grew within 5.7 h to an optical density of 2.0 (CDM = 0.7 g/L). The growth of the cells was slightly bi-phasic. During rapid glucose consumption and exponential growth, a rapid and significant formation of acetate was observed. After glucose consumption, the intermediately formed acetate was re-utilized with a rate equal to its production rate. The butyrate concentration in the media never exceeded background levels.

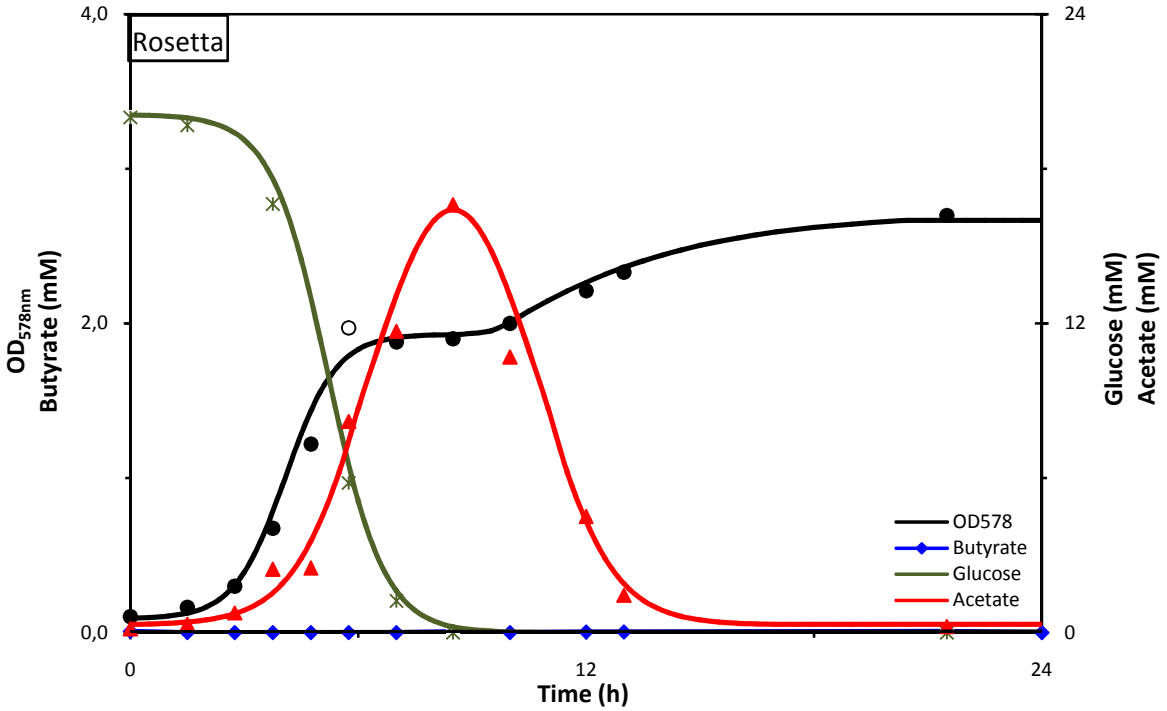


Fig.4.43.a: Growth curve, glucose consumption, butyrate and acetate concentrations in *E. coli* Rosetta culture. Note, all the curves were simulated using exponential equation and the value represented by an open symbol was taken from the simulation.

The recombinant strains grew significantly slower than the host did while the biomass production in these cultures was slightly reduced (0.46 and 0.5 g CDM/L at time point 5.7 h). In sharp contrast, the recombinant strains readily started to produce butyrate after induction of the foreign genes by anhydrotetracycline (Fig.4.43.b). It is rather obvious from the data depicted that the butyrate formation occurred on the expense of acetate. Its maximum concentration was significantly reduced in the cultures of recombinant strains.

It should be noted, that the introduction of butyrate fermentation pathway genes into *E. coli* was originally expected to promote not only butyrate formation from acetyl-CoA, but also to enable butyrate oxidation in the opposite direction. Thus, it must be considered remarkable that no butyrate oxidation was observed after glucose consumption by the recombinant strains at all, while acetate was oxidized to an extent, which is similar to the host (Fig.4.43.b).

While butyrate formation is shared by both strains and the final concentration of butyrate is even slightly higher in BUT_{AC} harboring the butyryl-CoA dehydrogenase from *C. acetobutylicum*, the kinetics of product formation differs significantly between both strains. It can be clearly seen in figure (4.43.b) that butyrate formation is readily induced by the addition of anhydrotetracycline in BUT_D while its formation occurs strongly delayed in BUT_{AC}.

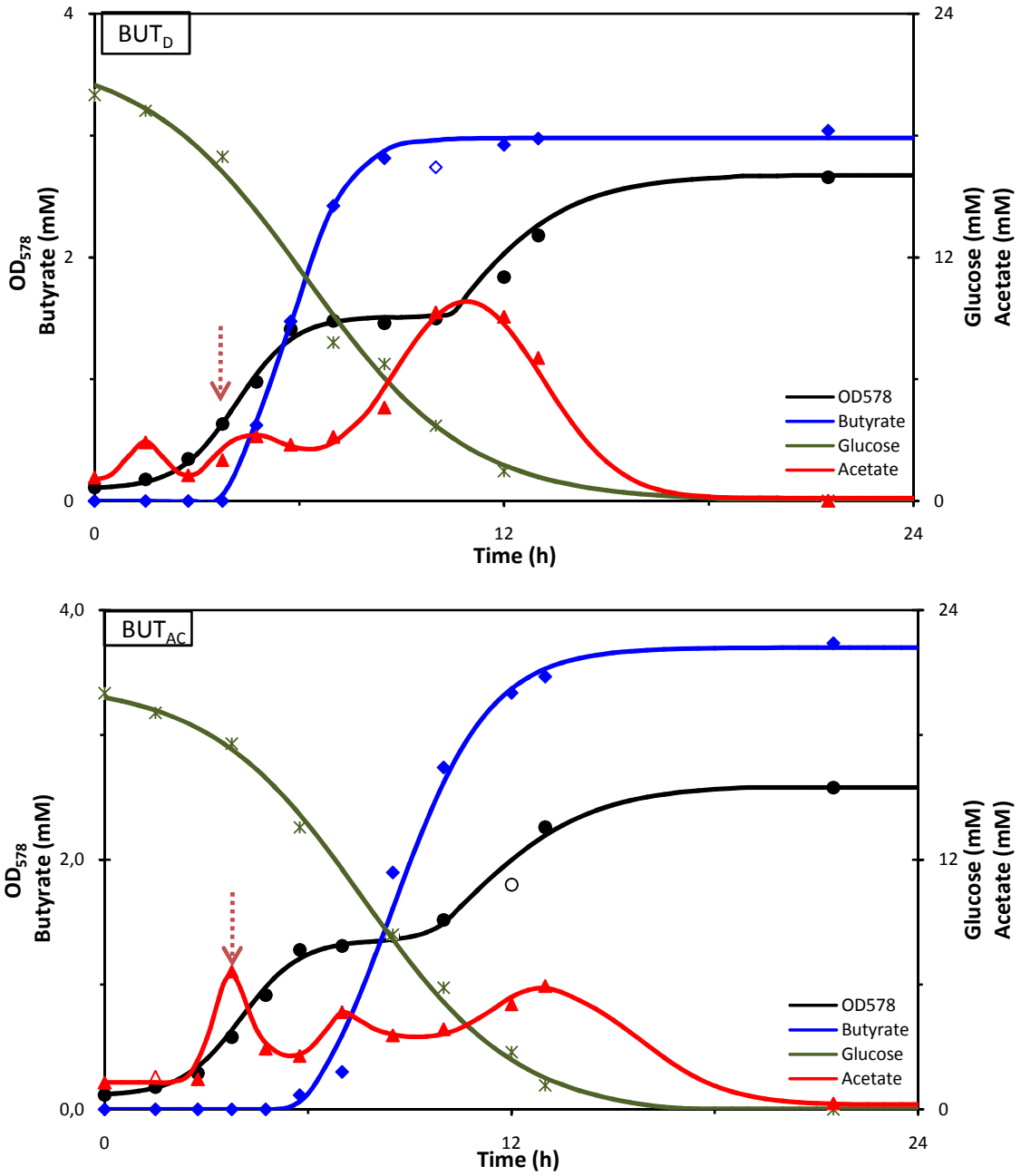


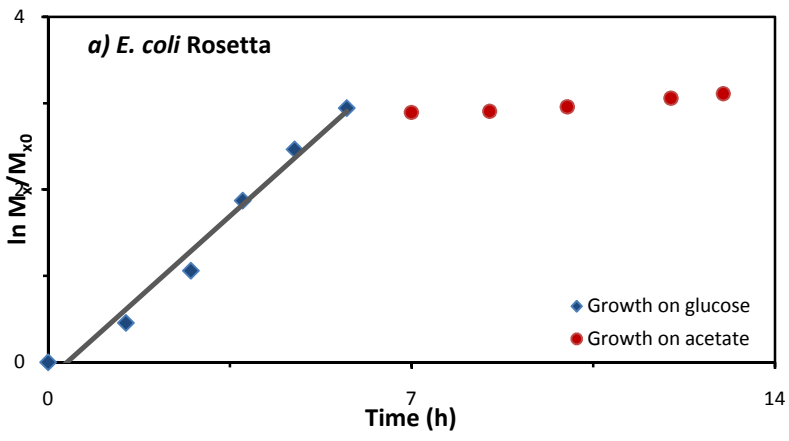
Fig.4.43.b: Growth curve, glucose consumption, butyrate and acetate concentrations in recombinant *E. coli* strains (BUT_D & BUT_{AC}) cultures. All the curves were simulated using exponential equation and values represented by open symbols were taken from the simulation. The induction-point is reflected by brown dotted arrow**.**

4.3.3. Models of butyrate fermentation

The kinetic model of batch fermentation provides information to understand cell mass and product generation, when the basic parameters of a fermentation process are known. Therefore, kinetic growth parameters, substrate utilization, product- and by-product formation must be known. In the following section, these parameters are calculated in order to provide information necessary to discuss the experimental findings in detail.

4.3.3.1.Biomass production

The growth of microorganisms in batch cultures occurs in a characteristic, multi-phasic manner: In the so-called lag phase the organisms must adapt to medium conditions. Subsequently, the microbes grow with maximum rate during exponential phase until the nutrient supply becomes limiting in the stationary phase. The maximum specific growth rate (μ^{max}) is a very characteristic kinetic parameter of growing organisms and it has a constant value for cultures growing under identical conditions. The μ^{max} is obtained in the exponential phase by plotting ln (M_X/M_{X0}) against time, where M_X is the C-mmol biomass at a given time point and M_{X0} the C-mmol biomass at the beginning of the experiment. The C-mmol biomass calculated from the equation (C-mmol= 1/25 mg CDM, where 25 is the molecular weight in grams for the biomass formula $C_1 H_{1.8} O_{0.5} N_{0.2}$). Applied on a non-disturbed growth situation, as can be assumed for the host cells, a straight line is observed with a slope that equals μ^{max} throughout substrate consumption (Fig.4.44). In contrast, any disturbances introduced (e.g. by induction of foreign genes) must definitely alter the appearance of this graph in a way that a stair-like plot is observed. As shown in figure (4.44), this is indeed the case, when the host cell behaviour is compared to the recombinant strains. While *E. coli* Rosetta grew upon glucose consumption with a rather constant μ^{max} of 0.54 h^{-1}, BUT_D, and BUT_{AC} strains exhibited only slightly reduced μ^{max} of 0.46 and 0.44 h^{-1} pre-induction (a fact that might be attributed to the need to sustain plasmid proliferation) the picture is significantly changed upon induction of the genes by anhydrotetracycline. Now, the energy demanding protein synthesis takes place and, moreover, significant part of the intracellular carbon flux is redirected into butyrate formation. In consequence, an almost complete intermediate arrest of cell growth is observed (Table 4.8). Noteworthy, this arrest is of transient nature, which suggests that the induction of the foreign genes by the light-sensitive inducer is not stable over time and might be decreased in later stages of the fermentation process. Indeed, it was observed that the specific activities for butyrate forming enzymes decreased at the end of the exponential growth phase but remained detectable even in stationary phase at reduced rates.



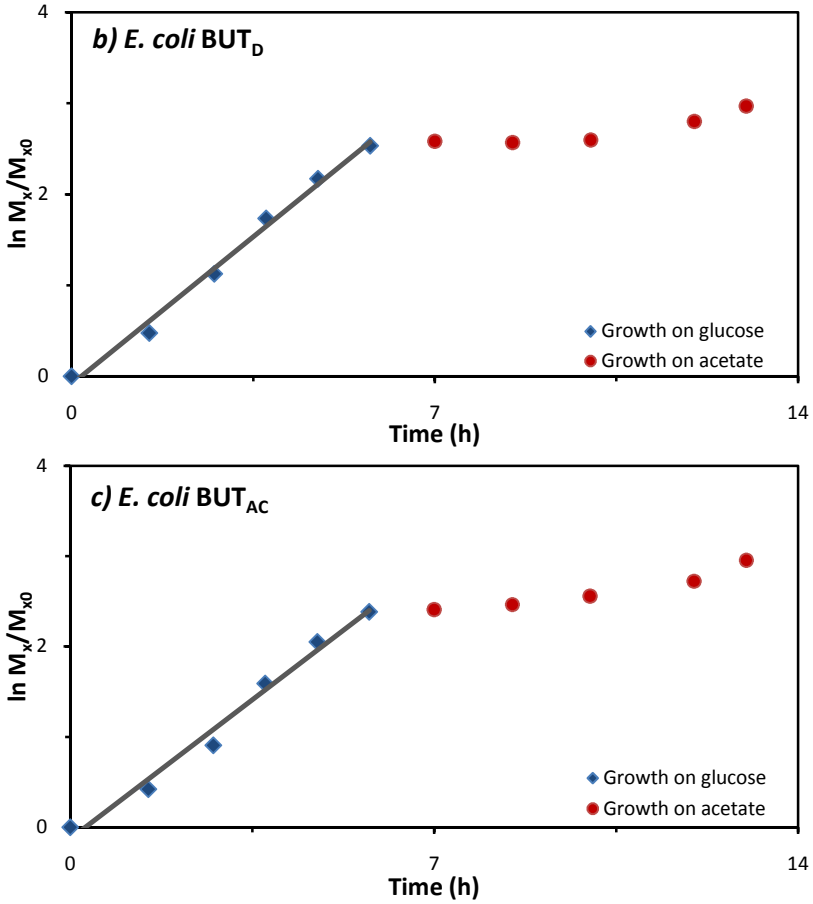


Fig.4.44: Maximum specific growth rate (μ^{max}) for *E. coli* Rosetta, BUT_D, & BUT_{AC} Strains, where M_X, the C-mmol biomass at a given time point; M_{X0}, the C-mmol biomass at zero time.

Table 4.8: Changes of specific growth rate (μ) over time.

interval time	Specific growth rate (μ) per hour		
	Rosetta	BUT_D	BUT_{AC}
0.00- 1.50	0.299	0.311	0.277
1.50- 2.75	0.469	0.505	0.380
2.75- 3.75	0.769	0.589	0.659
3.75- 4.75	0.568	0.427	0.453
4.75- 5.75	0.478	0.360	0.352
5.75- 7.00	0.000	0.039	0.000
7.00- 8.50	0.007	0.000	0.055
8.50- 10.0	0.014	0.018	0.060
10.0- 12.0	0.025	0.102	0.084
12.0- 13.0	0.029	0.169	0.227
13.0- 21.5	0.029	0.023	0.024

The Gray-shading cells shows the induction time.

4.3.3.2.Glucose utilization

The carbon source provided in a growth medium in batch cultures is generally used to generate metabolic energy for cell growth. It should be noted, however, that in particular under aerobic conditions, the majority of the carbon ends up in biomass. In fact, the total glucose consumption (Y_G) is actually linked to the amounts individually used for energy metabolism (catabolism, Y_C) and biomass production (anabolism, Y_A) and it holds that

$$Y_G = Y_C + Y_A$$

When the values entering the different pathways are known, it is possible to calculate the amount of ATP, which is formed in the energy metabolism. For *E. coli*, this value is close to 20 ATP per mole of glucose, and it is readily calculated that upon growth on glucose only about 20% of the glucose are used in the energy metabolism while about 80% end up in biomass.

The introduction of foreign genes, however, strongly alters this balance. Carbon which is irreversibly converted to butyrate significantly decreases the amount of ATP, which can be formed in average per mole of glucose (e.g. the reduction of 2 mole of acetyl-CoA to butyrate yields 1 mole of ATP, while the oxidation of two molecules of acetyl-CoA via the Krebs cycle yields 14 ATP in *E. coli*). Thus, it is very important to determine the fractional amounts of glucose utilized for different purposes in the recombinant and host strains. It should be noted, however, that only the value for total glucose consumption and biomass formation are readily assessable by measurements, while the values for product and by-product formations must be dissected by some theoretical considerations.

The conversion of glucose to biomass, $Yield_{(G/X)}$, is obtained in the exponential phase by plotting (M_{Gt}-M_{G0}) versus (M_{Xt}-M_{X0}) where M_{Gt} is the mmol glucose at time t; M_{G0}, the mmol glucose at the beginning of the experiments; M_X, the C-mmol biomass at a given time point; M_{X0}, the C-mmol biomass at the beginning of the experiments. A trendline is observed with a slope of $Y_{(G/X)}$ equal to q_G^{max}/μ^{max} (Fig.4.45), where q_G^{max} *is* the maximum specific glucose utilization (mmol glucose consumed/ h/ C-mmol biomass produced). The conversion of glucose to biomass for *E. coli* Rosetta, BUT_D, and BUT_{AC} strains were 0.53, 0.34, and 0.37 mmol glucose per producible C-mmol biomass, respectively. Additionally, (q_G^{max}) was of 0.29, 0.16 and 0.16 mmol glucose per hour per C-mmol biomass. It seems that the two butyrate strains have equal q_G^{max} with a significant decrease of about 45% compared to the control. This reflects a positive effect of the induction of foreign genes on glucose uptake and utilization, under conditions, where the flow of the metabolic flux through glycolysis is limiting for the utilization of glucose. In addition, the overflow of acetyl-CoA normally

observed is used to produce butyrate beside biomass rather than acetate, probably because it is more beneficial in terms of redox balance under condiditions where plenty of intermediates from TCA-cycle are shuttled into biosynthetic pathways.

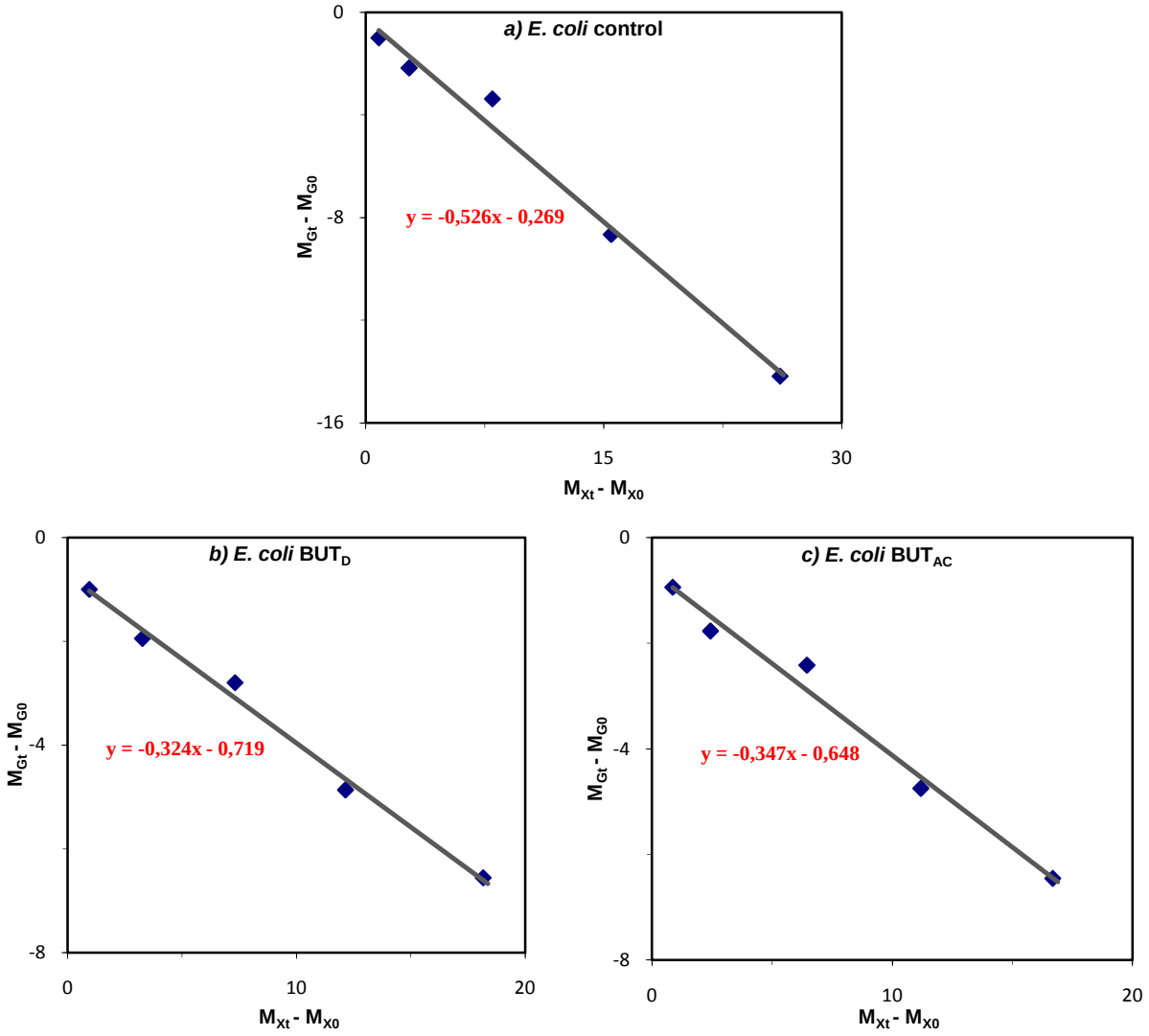


Fig.4.45: Maximum specific substrate utilization rate (q_G^{max}), where M_{Gt} is the mmol glucose at time t; M_{G0}, the mmol glucose at the beginning of the experiments; M_X, the C-mmol biomass at a given time point; M_{X0}, the C-mmol biomass at the beginning of the experiments.

In the control, the maximum consumption of glucose was observed at the end of the exponential growth phase (Fig.4.43.a). In contrast, a decrease in glucose consumption was monitored in butyrate producing strains (Fig.4.43.b) with conversion of glucose to biomass, acetate and butyrate. Thus, glucose uptake was related to biomass and butyrate production instead of biomass and acetate in control. The maximum productivity of biomass and butyrate was linked to a decline in glucose concentration in the culture. It is, therefore, likely that butyrate strains needed longer time to consume glucose as evident by its total consumption within 13 h in the recombinant strains as compared to 7 h in the control.

4.3.3.3.Butyric acid production.

As mentioned above, the host strain *E. coli* Rosetta produced only trace levels of butyrate (less than 0.03 mM). It was evident from the enzyme activities in cell free extracts (Table. 4.6) that *E. coli* Rosetta is not capable of producing significant amounts of butyric acid. In contrast, the modified strains harboring plasmids with the genes encoding butyrate-forming clostridial genes produce significant amounts of butyrate (Fig.4.43.b).

The maximum rate for butyrate production (R_B^{max}) in modified strains was determined by plotting the changes on butyrate production over time (Fig.4.46). No butyric acid was observed before the addition of the inducer anhydrotetracycline. *E. coli* BUT_D strain started to produce butyrate readily post-induction. The exponential production of butyrate was achieved for short period, from inducing point 3.75 to point 7.0 h, with butyrate concentration rising from 0.08 to 2.43 mM, respectively. Then, the production decreased due to the limitation of glucose in the culture. The maximum production of butyrate was about 3.04 mM (268 mg/l) after 13 h post- inoculation, when glucose-limitation in the culture occurred (Fig.4.43.b). It is worth noting that the butyrate production in this strain stopped fairly prior total glucose consumption and that a limitation of the cells in terms of enzyme activities can be excluded by the activity measurements shown in table (4.6).

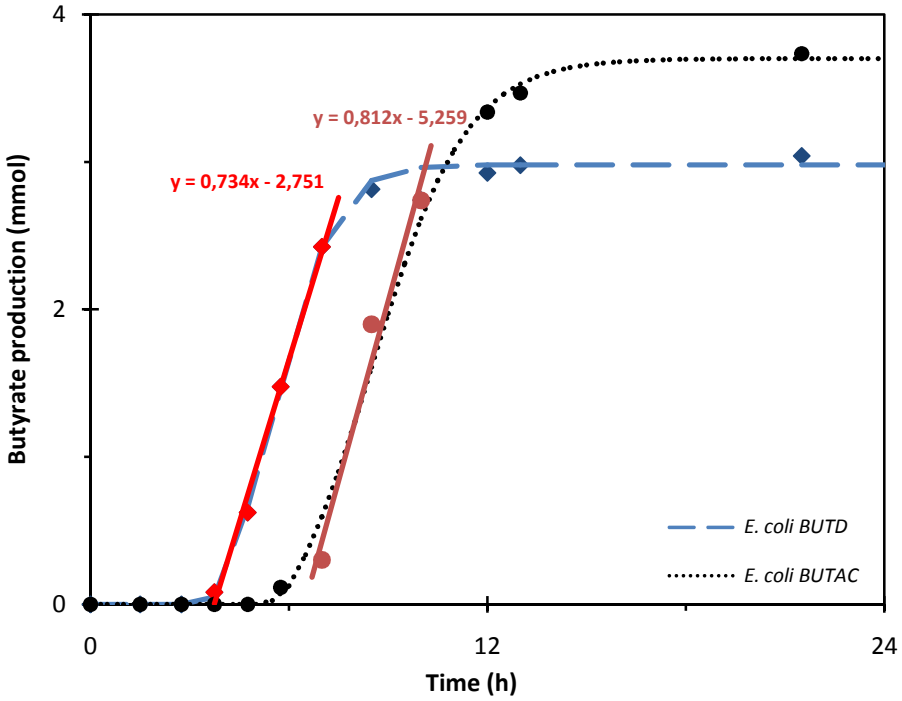


Fig.4.46: Butyric acid production over time in butyrate strains. The dashed blue line with a square mark (– –◆– –) refers to butyrate production rate in *E. coli* BUT_D strain, while the dotted black line with black cycle mark (·····●·····) refers to *E. coli* BUT_{AC}, while strain arrow showed the induction point.

On the other hand, the *E. coli* BUT_{Ac} strain started to produce butyrate only about two hours post-induction and exponential butyrate production started even later. The delayed butyrate production was accompanied by a sharp increase of acetate production (Fig.4.43.b). The data in figure (4.46) show that the maximum rates of butyrate production, obtained during exponential butyrate production phase, were 0.74 and 0.81 mmol butyrate per hour for *E. coli* BUT_D and BUT_{Ac} strains, respectively.

4.3.3.4.Acetic acid production

Acetic acid is a by-product during metabolic growth of *E. coli*. During the growth under aerobic conditions, the overflow of acetyl-CoA leads to acetate production to gain another ATP molecule instead of going into the citrate cycle, thereby increasing the redox potential in the cells. For that, the acetate production curve is very similar to the growth curve and increased with increasing cell growth and vice versa. In the control cells, exponential acetate production starts after about 2.75 h post-inoculation and a maximum of 16 mM acetate was observed after about 8.5 h (Fig.4.47). It is worth noting that after glucose consumption, the cells utilized acetate as an alternative carbon source with rates similar to its production rate and a smaller but significant growth was observed (Table 4.8).

In the modified strain *E. coli* BUT_D, the productivity of acetate was lower than in the control. Additionally, the maximum acetate production was reduced significantly to reach 9.3 mM acetate after 10 h post-inoculation (Fig.4.47). This reflects the redirection of parts of the metabolic fluxes towards butyrate, which can be attributed to the expression of butyrogenic genes. Notably, also the recombinant strains reutilized acetate readily after glucose consumption.

In contrast, the acetate production in *E. coli* BUT_{AC} resembled the control prior induction, where the maximum acetate concentration was 6.63 mM at time point 3.75 h (Fig.4.47). Suddenly after induction of foreign genes, here the acetate concentration decreased sharply. Notably, this decrease is not directly linked with butyrate formation, which occurred delayed. A significant increase in acetate concentrations was found to combine the first period of butyrate production (time between 5.75 to 7.0 h) to reach 4.68 mM, followed by a decrease in acetate production, which was linked with the exponential production of butyrate. Afterwards, the cells began to produce additional acetate, which was also readily consumed in the “stationary” phase.

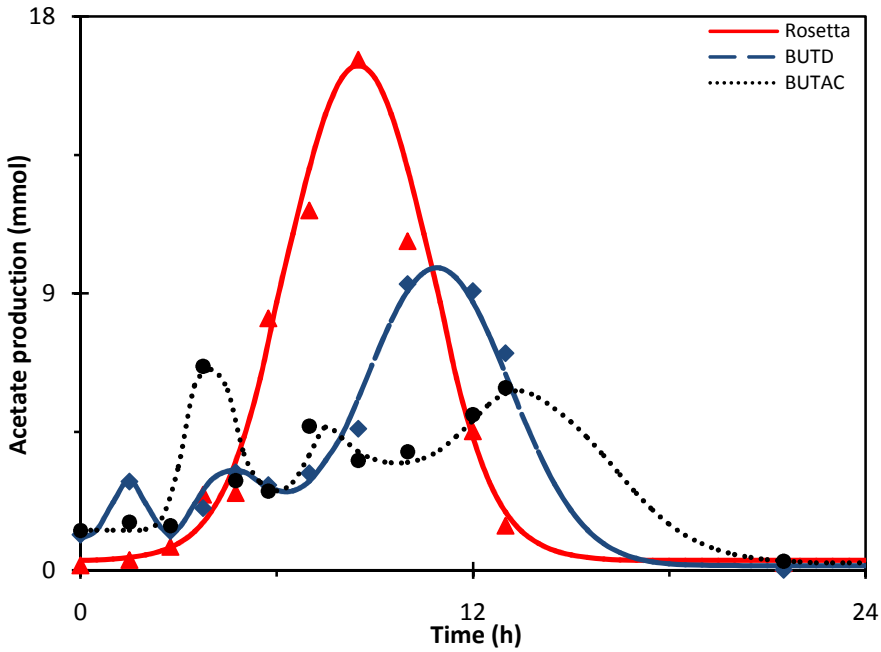


Fig.4.47: Acetate production over time in *E. coli* Rosetta, BUT_D and BUT_{AC} strains.

4.3.3.5.Operational yields

The data evaluation described so far enables the generation of time dependent plots of the operational yields of butyrate ($Y_{B/X}$) and ($Y_{B/G}$), from which the biomass-specific product formation (μ) can be obtained. Therein, $Y_{B/X}$ reflects the yield of mmol butyrate per C-mmol biomass produced while ($Y_{B/G}$) gives the yield of mmol butyrate per mmol of glucose consumed. When these data were plotted versus the incubation time (Fig.4.48), the differences between the two recombinant strains became much clearer. Rather obviously, in both strains the induction of foreign genes was accompanied by a significant but transient drop in the values for μ. More remarkably, butyrate formation was clearly linked with the exponential growth in cells harbouring the butyryl-CoA dehydrogenase from *C. difficile*, while its formation occurred predominantly in the stationary phase in cells harboring the corresponding genes from *C. acetobutylicum*. This fact becomes even more apparent, when the rate of butyrate formation was based on the glucose consumption as indicated by the $Yield_{B/G}$ values.

In the strain harboring *C. difficile* butyryl-CoA dehydrogenase, the maximum $Yield_{(B/X)}$ was 1.36 mmol butyrate per C-mmol biomass per L with a $Y_{B/G}$ of 0.503 mmol butyrate produced per 1 mmol glucose consumed. The theoretical conversion ratio of glucose to butyrate is 1:1. Thus, the stoichiometry of glucose consumption to butyrate formation by *E. coli* BUT_D was above half of the theoretical maximum during exponential growth.

On the other hand, in *E. coli* BUT_{AC} the butyrate $Yield_{(B/X)}$ was 1.43 mmol butyrate per C-mmol biomass, showing a glucose conversion rate of 0.535 mmol butyrate produced per mmol glucose consumed (Fig.4.49.b). Thus, also this strain produced butyrate with similar maximum efficiency, although kinetically delayed. Since the butyryl-CoA dehydrogenases encoded in the information carrying plasmids were the only difference between the recombinant strains, these observations might reflect different kinetic properties of butyryl-CoA dehydrogenase and its two electrons transfer subunits between *C. difficile* and *C. acetobutylicum in vivo*.

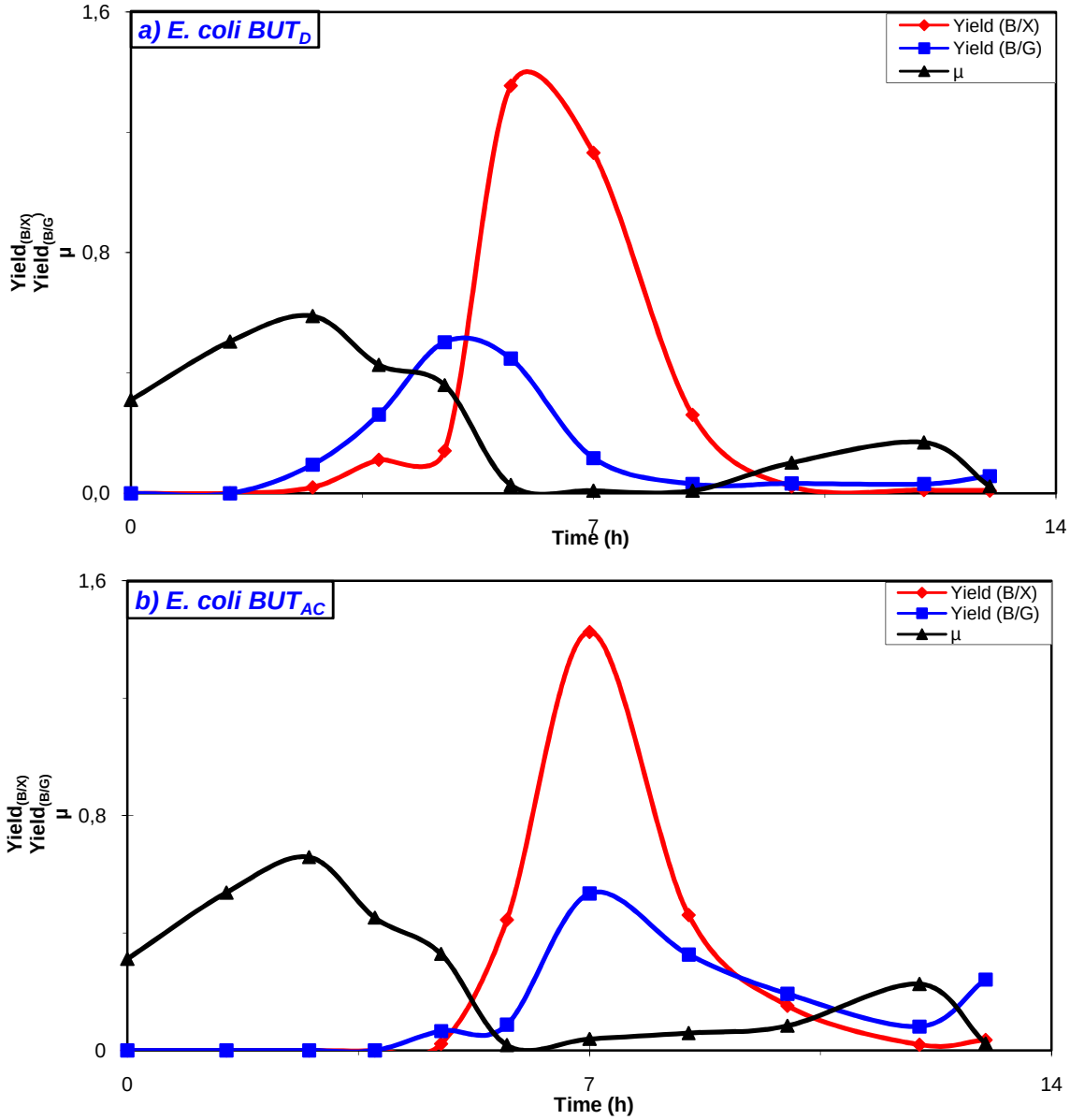


Fig.4.48: Changes of operational Yields and specific biomass production over incubation time for butyrate strains. Abbreviation in legend; $Yield_{(B/X)}$, yield of mmol butyrate per C-mmol biomass produced; $Yield_{(B/G)}$, yield of mmol butyric acid per mmol glucose consumed; μ, specific biomass production. Please note that, the (-) sign for $Yield_{(B/G)}$ which refer to the consumption of glucose is removed.

5. Discussion

The aim of this study was to elucidate a novel cloning and expression system for systematic applications in synthetic biology. Trial and error approaches in this context are rather costly in terms of time and money, and the development of structured routine protocols was supposed to be a reasonable approach.

As a model system, we have chosen the butyrate fermentation of *C. difficile*, which has not been studied in detail before. Here, the major question to address was whether or not the butyrate forming pathway in this organism involves a novel member of the recently discovered bi-forking butyryl-CoA dehydrogenases or not.

In order to succeed, individual enzymes were initially produced in *E. coli* in order to demonstrate their functional production followed by an initial characterization *in vitro*.

Successively, the enzyme encoding genes were assembled in synthetic operon structures to create artificial metabolic pathway modules, which could be expressed from a single plasmid and enabled the formation of the predicted product, butyrate. Thus, it has been demonstrated that crucial intermediates shared by different metabolic pathways (e.g. butyryl-CoA as an intermediate of butan-1-ol production or polyketide synthesis) can be formed using the information provided by a single plasmid. Given that *E. coli* can harbor multiple plasmids, a modular testing of various afferent (substrate delivering) and divergent (product forming) pathways is readily possible and might provide rapid access to the development of novel (synthetic) metabolic pathways in combinatorial approaches.

5.1. Protein purification and *in vitro* characterization

Using the Stargate™ cloning system, the seven enzymes in question from *C. difficile* as well as the control butyryl-CoA dehydrogenase from *C. acetobutylicum* were successfully produced in a functional state in *E. coli*. The initial characterization of individual enzymes was very much facilitated by the easiness to create enzyme variants with StrepII-tags for affinity purification: Thiolase, β-hydroxybutyryl-CoA dehydrogenase, crotonase, butyryl-CoA dehydrogenase/ETF complex, phosphate butyryltransferase and butyrate kinase were readily produced in an active state in the host and could be used in order to obtain reliable kinetic data for the enzymes in question in reasonable time. The results of these *in vitro* studies will be discussed for the individual enzymes in the following sections.

5.1.1. Thiolase

The first step in the metabolic pathway for butyrate formation in clostridial species is the condensation of two molecules of acetyl-CoA to yield acetoacetyl-CoA. This reaction is catalyzed by the enzyme acetyl-CoA-C-acetyl transferase (EC 2.3.9.1, further referred to as thiolase). Recombinant thiolase from *C. difficile* consists of 402 amino acid (the wild type has 391 amino acids) with a calculated molecular mass of 42.2 kDa. This is in agreement with the molecular weight of thiolase from different organisms such as *C. acetobutylicum* which was about 42 kDa **[19, 20]** and 44 kDa **[21]**, 41 kDa in *C. kluyveri* **[22]** and 41.5 kDa in *E. coli* **[110]**. Thiolase from *C. difficile* is a tetramer with a native molecular mass of 180±5 kDa (4*45 kDa) and it is closely related to thiolase from different bacteria, for which a tetrameric structure with native masses between 150-190 kDa have been reported**[21, 22, 111, 112]**.

The values observed for the V_{max} and K_{cat} of the recombinant enzyme from *C. difficile* were significantly higher than those previously reported for enzymes derived from other bacterial sources, while the K_M-values for the substrates fell well into the ranges previously described (Table 5.1).

Table (5.1): Comparison of the kinetic parameters of thiolase

Organism	$K_{M\ (\mu M)}$		V_{max} (U/mg)	K_{cat} (s^{-1})	Reference
	Acetoacetyl-CoA	CoA-SH			
C. difficile	**33**	**27**	**625**	**470**	**This work**
C. acetobutylicum	32	4.8	216	160	**[21]**
Bacillus subtilis	50	70	114	80	**[113]**

According to the mechanism, parallel straight lines were observed in reciprocal plots of initial velocity versus CoA-SH in the presence of fixed acetoacetyl-CoA concentrations. The parallel lines indicate that the kinetic mechanism for thiolase follows a Ping Pong pattern (Fig. 5.1), a behavior, which has been reported before in different thiolases from various sources **[21, 23, 114-116]**.

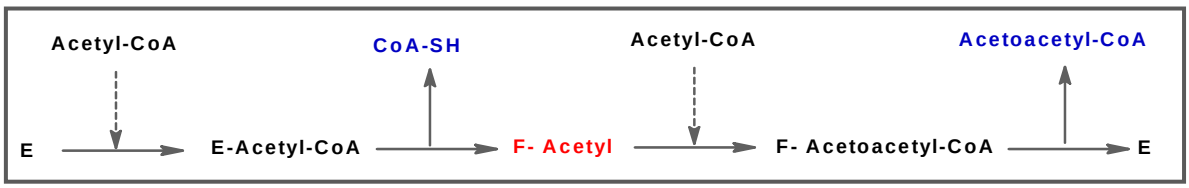


Fig.5.1: Simplified schematic view of the Ping Pong mechanism of thiolase.

5.1.2. β-Hydroxybutyryl-CoA dehydrogenase

Acetoacetyl-CoA formed by thiolase is then reduced with NADH to yield hydroxybutyryl-CoA. This reaction is catalyzed by β-hydroxybutyryl-CoA dehydrogenase (EC 1.1.1.157, further referred to as BHBD). The subunit mass for the recombinant BHBD was calculated to be 32 kDa and in agreement with the one from *C. acetobutylicum* (30.5 kDa **[25]**). It should be noted that *C. kluyveri* has two *hbd* genes with accession numbers CKL-0458 and CKL-2795. The encoded proteins consist of 282 and 319 amino acids with predicted molecular mass of 30.3 and 35.4 kDa, respectively. The native molecular mass for *C. difficile* BHBD is 144±3 kDa and suggests a tetrameric structure (4*33.5 kDa), which is in disagreement with BHBD from *C. kluyveri* that has been reported to form an octamer (8*26 kDa **[24]**).

The BHBD from *C. difficile* differs from the previously studied one in *C. kluyveri* (Table 5.2) in that the former is NADH-dependent in contrast to the latter, which has been shown to be NADPH-dependent **[24]**. The authors observed a second enzyme in the cell lysates, which was, however, NADH-dependent. These results suggest that *C. difficile* BHBD is similar to enzymes in the butyrate forming pathways of other clostridia. It indicates that NADH is naturally involved in fermentative butyrate synthesis as reported by ***Hartmanis and Gatenbeck*** for BHBD from *C. acetobutylicum* **[27]** and that the NADPH-dependent enzyme in *C. kluyveri* might be involved in hitherto unknown biosynthetic pathways.

Table (5.2): Comparison of the kinetic parameters of BHBD

Organism	$K_{M\,(\mu M)}$			V_{max} (U/mg)	K_{cat} (s^{-1})	Reference
	Acetoacetyl-CoA	NADH	NADPH			
C. difficile	**17**	**210**	**--**	**54**	**30**	**This work**
C. kluyveri	50	---	70	290- 450	195	**[22, 24]**

BHBD from *C. difficile* was substrate inhibited by acetoacetyl-CoA. It has been previously suggested that the enolate–form of the substrate is a potent competitive inhibitor, which resembles both crotonyl-CoA and β-hydroxybutyryl-CoA **[30]**.

The pH-dependent activity profile of BHBD from *C. difficile* resembles the one previously described for the *C. kluyveri* enzyme, which has an optimum at pH 6.5 with an activity pH range covered from 5.5 to 7.0 with about 90% relative activity **[24].** BHBD from *C. saccharobutylicum* is known to exhibit higher activity during solvent production at pH 5.5 **[117]**.

The kinetic mechanism of the reaction catalyzed by the bi-reactant enzyme BHBD has been studied extensively. The enzyme can be largely classified into two classes, following either a sequential or a Ping Pong mechanism. Intersecting straight lines observed in the double reciprocal plots of the kinetic data of BHBD from *C. difficile* suggest that the enzyme follows a sequential, but not a Ping Pong mode. Since converting initial velocity patterns are characteristically for the ordered *bi-bi* and the random *bi-bi* mechanism, the reaction of BHBD follows one of these rapid bi-reactant mechanisms. The ordered addition of substrates is well established for many dehydrogenases: *D*-hydroxyisovalerate dehydrogenase from *Fusarium sambucinum* **[118]**, (*S*)-lactate dehydrogenase from *Plasmodium falciparum* **[119]** and 3-oxoacyl-ACP reductase from *Plasmodium falciparum* **[120]**are known to bind first the nicotinamide cofactor, followed by the binding of the substrate. In contrast, other dehydrogenases bind the substrate first. This binding mode has been shown for UDP-glucose dehydrogenase (EC 1.1.1.22) from beef liver **[121]**, yeast aldehyde dehydrogenase **[122]**, and histidinol dehydrogenase (EC 1.1.1.23) from *Salmonella typhimurium* **[123]**. BHBD also shows substrate inhibition by acetoacetyl-CoA and its inhibition was not observed with the coenzyme (NADH). This may reflect that NADH indeed binds first to the enzyme without affecting the binding of acetoacetyl-CoA. In turn, binding of acetoacetyl-CoA to the free enzyme inhibits the enzyme activity, probably decreasing the affinity of the enzyme-substrate complex for the Co-enzyme. This property suggests that only one substrate (NADH) binds to the free enzyme, while the second substrate is bound by the binary complex, and binding of acetoacetyl-CoA forms an inhibitory complex not capable of proceeding the catalytic cycle. Therefore, the reaction mechanism of BHBD likely follows an ordered *bi-bi* mechanisms with NADH as leader substrate (Fig. 5.2).

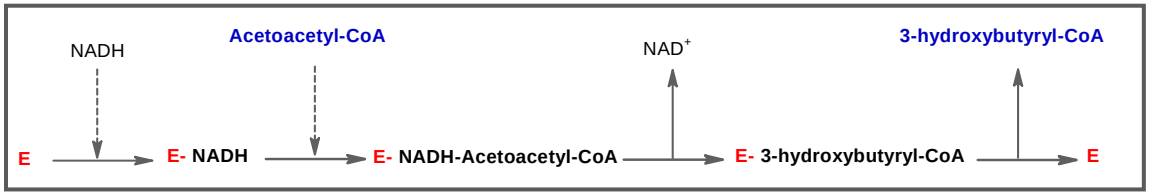


Fig.5.2: Schematic representation of the ordered *bi-bi* mechanism of BHBD.

5.1.3. Crotonase

The present study describes the purification and partial characterization of crotonase from *C. difficile*, an enzyme catalyzing the reversible *syn*-hydration of crotonyl-CoA. The subunit molecular mass was about 29.7 kDa as judged by SDS-PAGE and calculated from the amino acid sequence (276 amino acids). Molecular masses of 40 kDa has been described for

crotonase from *C. acetobutylicum* **[32],** but the molecular mass from amino acid sequence derived calculations (261 amino acid) is only 28.2 kDa **[25]**. Crotonase from *C. difficile* was a tetramer (4*29.7 kDa) with a native molecular mass of 125±3 kDa. This structure is in agreement with previous findings**[32]**. The oligomeric structure of *C. difficile* crotonase is also reflected by the observation, that the StrepII-tag fused to the recombinant variants strongly affected enzyme activities (table 5.3).

Table (5.3): Comparison of the kinetic parameters of crotonase

Organism	$K_{M\,(\mu M)}$ Crotonyl-CoA	V_{max} (U/mg)	K_{cat} (s^{-1})	Reference
C. difficile	**52**	**22,470**	**11,120**	**This work**
C. acetobutylicum	30	6,500	4,330	**[32]**

5.1.4. Butyryl-CoA dehydrogenase

The central question to be answered in this work was whether or not *Clostridium difficile* possesses a butyryl-CoA dehydrogenase, which belongs to the recently discovered sub-class of enzymes that couple NADH oxidation with the thermodynamically unfavourable reduction of ferredoxin **[33]**. Since validation of such an enzyme in the organism enables electron transport phosphorylation using the RNF-complex **[49]**, its detection is of general importance to understand the remarkable metabolic versatility previously considered for this emerging human pathogen **[10, 11]**.

The data presented here demonstrate that the butyryl-CoA dehydrogenase from *C. difficile* belongs to the recently described novel sub-family of bi-forking enzymes capable to couple the crotonyl-CoA reduction with the endergonic NADH-driven reduction of ferredoxin.

Butyryl-CoA dehydrogenases form a functional complex with electron transferring flavoproteins (ETF). In the crystal structure of human ETF, FAD and AMP are found as prosthetic groups **[124]**. The *Megasphaera elsdenii* enzyme is the only well-studied and structurally characterized bacterial BCD/ETF complex **[17]**. Here, the ETF contains 2 mol of FAD per heterodimer but no AMP; one FAD is covalently bound to the protein, whereas the other appears to readily dissociate from the enzyme and has unusual spectral properties **[125]**.

It has been previously shown that function of butyryl-CoA/ETF complexes recombinantly produced in *E. coli* is strongly affected by the position of affinity tags attached

to the nature polypeptides by the means of genetic manipulations for facile purification **[107]**. These data were previously obtained by ***Regina Fischer*** and the results are crucial for this work, too **[107]**. While all the recombinant enzyme complexes transferred butyryl-CoA dehydrogenase activity into the host, which was readily measurable in cell lysates, most of the recombinant enzymes were either inactive or significantly impaired in their catalytic activities after purification. Only the enzyme variant in which the StrepII-tag for affinity purification was fused to the C-terminus of the EtfA-subunit was fully active, while in other variants the activity was strongly reduced or virtually absent, due to complex decay upon purification. Thus, all data presented herein refer to this particular enzyme variant.

The recombinant BCD from *C. difficile* was purified under oxic conditions and the activity was initially measured following the oxidation of butyryl-CoA with ferricenium hexaflourophosphate, which serve as an artificial, strong oxidizing electron acceptor **[101]**. Like other clostridial acyl-dehydrogenases, the enzyme consists of three subunits: α-subunit (butyryl-CoA dehydrogenase, 42.3 kDa), a large ETF β-subunit (38.7 kDa) and a small ETF γ-subunit (29 kDa) encoded by *bcd*, *etfA* and *etfB*, respectively. A stoichiometry of 1:1:1 and 1:2:1 was found for the three subunits (αβγ) from *C. difficile* and *C. acetobutylicum* produced in *E. coli*, respectively. For propionyl-CoA dehydrogenase isolated from *C. propionicum* a stoichiometry of 2:1:1 has been reported **[126],** as well as butyryl-CoA dehydrogenase purified from *C. kluyveri* **[33]**. This suggests that the enzymes from various sources differ in their composition.

The butyryl-CoA dehydrogenase from *C. difficile* is the first clostridial enzyme, which has been produced in *E. coli* and was purified to homogeneity prior to enzyme measurements. Previously, ***Inui*** and co-workers cloned the *crt-bcd-etfB-etfA-hbd* operon (4.8 kb) from *C. acetobutylicum* ATCC 824, expressed it in *E. coli*, measured the activity in the cell free extract and demonstrated that the ETF-subunits are essential for activity **[57]**. In line with its essential requirement, ETF is tightly bound in some dehydrogenase enzymes **[33, 126, 127]**. This holds also true for the enzyme from *C. difficile*.

Li and co-workers first purified the enzyme from *C. kluyveri* under anoxic conditions, which is the prototype of a bi-forking butyryl-CoA dehydrogenase, and found that this enzyme is absolutely ferredoxin-dependent **[33]**. They unquestionably demonstrated that reduced ferredoxin is a second product of the enzyme catalyzed reaction, which was quantified by means of hydrogenase-mediated H_2-release. On the other hand, ***Herrmann*** reported that FAD can at least partially substitute for ferredoxin in the enzyme from *C.*

tetanomorphum **[127]**. In line with the latter observation, we found that *C. difficile* BCD can also use free FAD as alternative electron acceptor. In this work, we developed a novel assay for BCD, in which oxidized ferredoxin is regenerated under anoxic conditions by soluble Ferredoxin:NADP oxidoreductase from *E. coli* on the expense of provided $NADP^+$ (Eq. 4.6). Under anoxic conditions and with limited FAD concentrations, NADH-driven crotonyl-CoA reduction was strictly dependent on the co-reduction of ferredoxin.

It should be noted, however, that the recombinant enzyme from *C. difficile* was rather oxygen stable and showed significant activity under oxic conditions, too. This is in particular true for the physiological direction of the catalyzed reaction, the reduction of crotonyl-CoA with NADH. The enzyme complex from *C. acetobutylicum*, which was analyzed as a control, was oxygen-sensitive and lost activity when stored under air. The *C. difficile* enzyme retained activity for several days. Freshly prepared enzymes from both organisms yielded readily measurable activity for the ferricenium hexafluorophosphate-dependent oxidation of butyryl-CoA. Both enzymes exhibited diaphorase activities with NADH, but only the enzyme from *C. difficile* was able to reduce crotonyl-CoA with NADH under air. In sharp contrast to the enzyme from *C. acetobutylicum*, which was entirely inactive with crotonyl-CoA and NADH under air when the release of butyryl-CoA is concerned, the enzyme from *C. difficile* carried out the reduction of crotonyl-CoA. For the latter enzyme, a crotonyl-CoA-dependent oxidation of NADH was readily detected by increased consumption of NADH, and the predicted product, butyryl-CoA, was unquestionably detected by MALDI-TOF-MS. It should be noted that the oxygen sensitivity and the failure to measure the enzyme in the crotonyl-CoA reducing direction have been previously noted for the *C. acetobutylicum* enzyme **[27, 57, 59, 66]**. Thus, the enzyme from *C. difficile* is rather unique in that this enzyme can utilize O_2 as an alternative oxidant, which replaces the natural substrate (oxidized ferredoxin) under oxic conditions.

In a meaningful manner, the catalytic properties of the butyryl-CoA dehydrogenase from *C. difficile* can be compared only to the bi-forking enzymes previously described in *C. kluyveri* and *C. tetanomorphum*, although its ability to oxidize butyryl-CoA with ferricenium is shared with other enzymes. Regarding to the kinetics constants, butyryl-CoA dehydrogenase from *C. difficile* fell well in the range of values previously observed for enzymes from other sources (Table 5.4). The only significant difference is the K_M-value for NADH, which is almost 10-times higher for the *C. difficile* enzyme than previously reported for BCD from *C. kluyveri*.

Table (5.4): Range of kinetic constants for Bcd/ETF complexes from different sources.

Substrates	$K_{M\ (\mu M)}$	V_{max} *(U/mg)*	K_{cat} (s^{-1})	References
Butyryl-CoA	6-16	1.3- 331	2-237	**[127-129]**
	10	**15**	**28**	**This work**
Crotonyl-CoA	4	18	31	**[127]**
	2.5	**10**	**19**	**This work**
Ferredoxin	10	5-14	----	**[33]**
	2	**20**	**37**	**This work**
NADH	18	12	21	**[127]**
	145	**19**	**36**	**This work**

The values given were taken from the indicated references and the values for *C. difficile* BCD are show in bold numbers.

5.1.5. Butyrate kinase

Butyrate kinase (BUK) phosphorylates ADP on the expense of various acyl-phosphates. BUK from *C. difficile* has a molecular mass of 42.6 kDa, which is similar to the values reported for enzymes from *C. acetobutylicum* (39 kDa, **[45]** and 43 kDa for its isozyme, **[43])**. The enzyme from *C. difficile* is a dimer of two identical subunits with a native mass of 89±3 kDa like previously reported for the BUKs of *C. acetobutylicum*, were native masses of 80 kDa (2*39 kDa) for BUK **[45]** and 85 kDa (2*43 kDa) for BUK-isozyme have been reported **[43]**.

It is worth noting that in this study the activity of butyrate kinase was measured for the first time in its physiological, butyrate-forming direction using glucokinase and glucose-6-phosphate dehydrogenase to detect ATP synthesis. This method is more sensitive and accurate than the method of Rose, which measures activity in the reverse direction using a very small extinction co-efficient of ε_{540nm}= 0.169 $mM^{-1}cm^{-1}$ **[130]**. A reducing agent such as 1 mM DTT was found to be necessary for optimum activity during purification and storage of the enzyme. ***Hartmanis*** investigated the effect of treatment with sulfhydryl-modifying reagents and suggested that there is no involvement of sulfhydryl groups at the catalytic site **[45]**. A possible explanation for the requirement of a reducing agent for enzyme stability might be that the 4 cysteines per enzyme molecule upon oxidation form intra- or inter-chain disulfides and, thereby, alter the conformation of the enzyme leading to a sub-optimum activity **[45]**.

The Lineweaver-Burk plots of butyryl-phosphate concentrations in the presence of fixed ADP concentrations versus initially velocity presented straight lines which intersect on the x-axis. This indicates formation of a ternary complex of enzyme-butyrylphosphate-ADP in either a random or an ordered *bi-bi* mechanism.

Table (5.5): Comparison of the kinetic parameters of BUK

Organisms	K_M (µM)				V_{max} (U/mg)	K_{cat} (s^{-1})	References
	Butyryl-P	ADP	Butyrate	ATP			
C. difficile	**470**	**70**	**----**	**----**	**345**	**245**	**This work**
C. acetobutylicum BUK_I	-----	----	$14*10^3$	$1.4*10^3$	420	273	**[45]**
BUK_{II}			620	----	165	118	**[43]**
C. tetanomorphum	-----	----	$10.8*10^3$	$5.06*10^3$	65	-----	**[46]**
C. butyricum	----	----	$20*10^3$	$1.4*10^3$	163	-----	**[47]**

5.1.6. Phosphate butyryl-transferase

The release of butyrate from butyryl-CoA is coupled with energy conservation by substrate level phosphorylation. In the metabolism of anaerobes, the coupling of ATP regeneration is either achieved directly, employing phosphate butyryl-transferase and butyrate kinase, or indirectly, using a butyrate CoA-transferase for acetyl-CoA synthesis followed by ATP-synthesis aided by phosphate acetyl-transferase and acetate kinase. The investigation presented here focused on the first way, employing phosphate butyryl-transferase and butyrate kinase.

Phosphate butyryl-transferase (PBT) catalyzes the transfer of a butyryl-residue from coenzyme A to phosphate. PBT and related enzymes are widespread in microbes and capable of producing acyl-phosphates including acetyl- **[131]**, propionyl- **[132]**, butyryl- **[42]**, carbamayl- and aspartyl-phosphates, which are either employed in energy conservation or provide essential metabolic building blocks. PBT has been partially purified from *C. butyricum* and was also detected in several other clostridial species **[42].** It has been isolated form *C. acetobutylicum* ATCC 824 in order to establish its physical and kinetic properties **[40]**.

The molecular mass of the N-terminally-tagged PBT (34.2 kDa) from *C. difficile* falls well in the range of clostridial PBTs, which exhibit molecular masses of 33 kDa in *C. butyricum* **[39]** and 31 kDa in *C. acetobutylicum* **[40, 41]**.

Comparison of the kinetic parameters of PBT is hampered by the fact, that previously studied enzymes were measured with DTNB**[133]** in order to detect the release of free CoA (table 5.6). The enzyme from *C. difficile* was measured using a novel coupled assay, in which the butyrate kinase-mediated ATP-formation was detected with glucokinase and glucose-6-phosphate dehydrogenase. Although this hampers the comparison of the enzyme with its counterparts in other organisms, it is closer to the *in situ* metabolic context, and might, therefore, provide data more relevant in the assessment of the enzymes function *in vivo*.

Table (5.6): Comparison of the kinetic parameters of PBT

Organisms	$K_{M\,(\mu M)}$ Butyryl-CoA	V_{max} (U/mg)	K_{cat} (s^{-1})	References
C. difficile	**72**	**257**	**151**	**This work**
C. acetobutylicum	110	1,380	713	**[40]**
C. beijerinckii	40	1,560	858	**[39]**

The kinetics of PBT followed a classic random *bi- bi* mechanism (Fig. 5.3). Enzymes from other sources exhibited broad substrate specificity and utilized various substrates including propionyl-CoA, valeryl-CoA, isovaleryl-CoA, and isobutyryl-CoA as described for PBT from *C. acetobutylicum* **[40],** while acetyl-CoA (relative specific activity against butyryl-CoA 35%) and acetoacetyl-CoA (relative specific activity 83%) were additional substrates for PBT from *C. butyricum* **[39]**. Likewise, PBT from *C. difficile* exhibited relative specific activities of 5, 26 and 33% for acetyl-CoA, acetoacetyl-CoA and propionyl-CoA, respectively, when assayed with Ellman's reagent**[133]**. Likewise, the relative activities with alternative substrates were 4, 18 and 23% in the coupled assay, suggesting that also butyrate kinase from *C. difficile* exhibits rather broad substrate specificity.

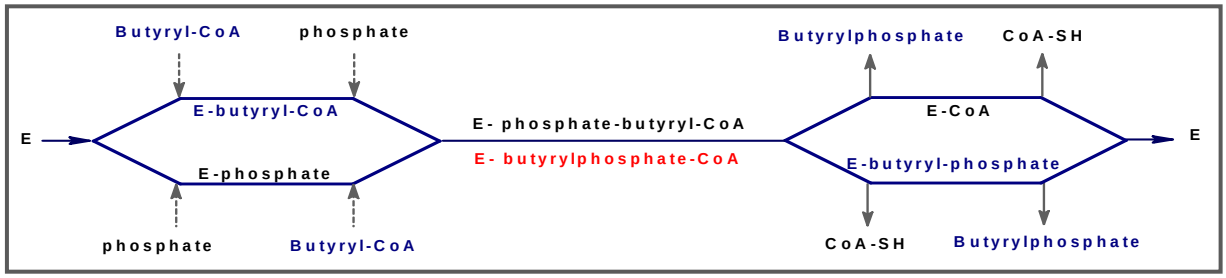


Fig.5.3: Simplified schematic of the random *bi-bi* mechanism of phosphate butyryl-transferase.

5.2. Butyrate production by the recombinant strains

The data presented thus far clearly demonstrated that the *C. difficile*-derived genes encoding enzymes of the butyrate forming pathway are readily expressed in *E. coli* and yielded functional recombinant enzymes. In the next step, individual genes were combined in single expression vectors, which harbored all the genes required for butyrate formation. In order to elucidate the different properties of the "conventional" butyryl-CoA dehydrogenase from *C. acetobutylicum* and the bi-forking one from *C. difficile* in a physiological context, two plasmids (*pASG-BUT$_{AC}$* and *pASG-BUT$_{D}$*) were created, which exclusively differed in the origin of the genes encoding the BCD complex. All other genes were from *C. difficile*. Transfer of these genes into *E. coli*-Rosetta pLysE created two recombinant strains, *E. coli* BUT$_{AC}$ and *E. coli* BUT$_{D,}$ respectively. Both strains and the host strain were then grown in minimal medium with glucose as sole source of energy and carbon in order to establish growth related fermentation parameters including specific growth rates, substrate consumption and product formation rates.

As expected, only the recombinant strains, *E. coli* BUT$_{AC}$ and BUT$_{D}$, were capable of forming butyrate and maximum values of 3.1 and 3.7 mM were obtained with a maximum yield 0.54 and 0.51 mmol butyrate per mmol glucose consumed. Likewise, the maximum production rates were 0.81 and 0.74 mmol butyrate per hour per C-mmol biomass and appeared rather similar. These data reflect that there is no significant difference between both strains in the maximum yield or the production rate. Likewise, the effects of the foreign gene expression on growth rate, glucose consumption and acetate formation were rather similar in the recombinant strains.

While the net results between both strains did not differ significantly, the kinetics of butyrate formation were rather different. While *E. coli* BUT$_{D}$ readily started to produce butyrate upon induction of the foreign genes, *E. coli* BUT$_{AC}$ butyrate formation was significantly delayed and occurred at high rates only at comparable high cell densities. This observation is in line with the oxygen sensitivity of the latter enzyme and has been previously reported by *Inui et al.* **[57]**, who found that butan-1-ol production in recombinant *E. coli* harboring the *C. acetobutylicum* solventogenic genes was not observed in actively growing cells but in harvested cells incubated under anoxic conditions. Notably, this behavior was attributed to the oxygen sensitivity of BCD also by these authors.

5.2.1. Effects of the foreign genes and their regulatory properties in *E. coli*

In order to address *in vivo* effects of foreign genes in *E. coli*, their influences in the normal carbon flux and their impact in the regulation of metabolic pathways of the host must be known. Acetyl-CoA is a common and central intermediate of glucose oxidation in *E. coli* under oxic and anoxic conditions **[77, 134-137]**. Under oxic conditions, mainly two major reactions compete for acetyl-CoA in the energy metabolism of the organism: The available pool of acetyl-CoA is predominantly shuttled via citrate into the Krebs cycle and finally oxidized to CO_2 and H_2O in the respiratory chain **[74, 138-142]**. It is, however, long known that significant portions of acetyl-CoA in aerobically, exponentially grown *E. coli* are redirected to substrate level phosphorylation as readily evident by the large concentration of acetate, which accumulates in cultures growing with glucose **[143-146]**. Likewise, it is well established that the intermediately formed acetate is readily end oxidized when glucose is no longer available as energy source **[78, 89, 90, 147]**. The host strain *E. coli* Rosetta pLysE nicely showed this behavior. In sharp contrast, butyrate once formed was not at all reutilized by the recombinant strains.

E. coli produces acetic acid as an extracellular by-product of aerobic fermentation and this exists in the ionic form at the neutral pH. The rate of acetate formation is directly related to the cell's growth rate, oxygen consumption rate, and the rate of glucose uptake **[139, 143, 148, 149]**. In the common operational mode of a fed-batch process, the growth rate of the culture is determined by the feeding rate of the limiting nutrient: specifically, *E. coli* generates acetate when glucose is the limiting nutrient and the cells grow above a threshold growth rate **[78, 148, 150]**. Despite the fully aerobic conditions, acetate formation is not related to the availability of oxygen, but rather relates to the rate of oxygen consumption. Regardless of the oxygen supply to the bioreactor, *E. coli* can only consume oxygen up to a maximum rate **[151]**. This plateau in oxygen consumption corresponds with the threshold growth rate at which acetate formation begins **[148]**. The value of the threshold growth rate depends on the strain but has been reported in defined media to be in the range of 0.14– 0.17 h^{-1} for fed-batch processes **[152]** and 0.35–0.48 h^{-1} for chemostats or accelerostats **[153, 154]**. The specific growth rate for the *E. coli* Rosetta strain, which was used in this work as control was 0.54 h^{-1} in M_9 medium. There are two explanations for the relationship between acetate accumulation in the culture and oxygen limitation. Firstly, the respiratory system, where NADH is reoxidized, has a limited capacity. As flux to the TCA cycle results in NADH production and flux to acetate does not, redirection of acetyl-CoA flux to acetate would be necessary to avoid accumulation of NADH when the respiration saturates **[8, 11]**. A second explanation is that

the TCA cycle has a limited capacity and that its limitation is reached before that of the respiration. When the TCA cycle saturates, increasing glucose uptake will redirect acetyl-CoA to form acetate. In this case, NADH production and respiration can increase further until the maximum respiration capacity or the maximum glucose uptake is reached **[155]**.

In the recombinant strains, the cells reach the maximum specific growth rate before the cells were induced for recombinant protein production. Indeed, initial acetate accumulation was likewise observed in these cultures. Foreign gene expression, however, caused a shift in the metabolic behavior of the cells and rather than acetate, butyrate was produced. At the same time, cells consumed more energy to produce recombinant protein as reflected by a growth arrest. Furthermore, recombinant proteins redirected the overflow of acetyl-CoA into the butyrate cycle as readily evident by product formation. ***Delisa*** *et al.* reported that the expression of a recombinant protein can significantly change fermentation process behavior, even when no intermediates are directly redirected **[82]**. ***Koh*** *et al.* reported that acetate becomes more inhibitory to recombinant protein producing cells than in the wild-type **[156]** and the threshold growth rate for the onset of acetate formation is lower for cells that are induced to generate a recombinant protein **[157]**. All these effects might be beneficial in terms of butyrate formation by the recombinant strains. Thus, the observation that *E. coli* BUT_{AC} showed a delayed butyrate production as compared with *E. coli* BUT_D might be best explained by the oxygen sensitivity of the *C. acetobutylicum* BCD **[57]**.

Introduction of expression plasmids harboring the genes of butyrate forming enzymes from clostridia introduce another competing metabolic pathway. Here, acetyl-CoA is partially captured by thiolase and shuttled into a reductive pathway, which finally yields butyrate. While this novel pathway increases the ability of the cell to reoxidize NADH formed in glycolysis and pyruvate dehydrogenase reaction, it decreases the ATP yield, which can be obtained by substrate level phosphorylation employing phosphate transacetylase and acetate kinase. Compared to butyric acid formation, formation of acetic acid yields twice as much ATP per mole of acetyl-CoA by substrate phosphorylation. Assuming that about 2 mole of ATP are generated by reoxidation of NADH within the truncated respiratory chain of *E. coli*, the ATP yield is further reduced to below half of that in the normal situation, when 2 out of 4 NADH formed in glycolysis and pyruvate oxidation are oxidized by butyrogenesis. A similar situation is given in fermentative growing clostridia, where the necessity to balance redox equilibrium by butyrate formation enforces the reduction of ATP yield, which could be obtained from acetyl-CoA. Therefore, the organism's capacity to regulate butyrogenesis and thus the ratio of acetate and butyrate is also vital to control the yield of ATP in anaerobes. In

this respect, it is worth noting that both citrate synthases and thiolases are known to be regulated by ATP as well as by the shared product CoA-SH and it has been suggested that both regulators compete with acetyl-CoA for the active site **[21, 158, 159]**.

In the thiolytic direction, CoA-SH has been demonstrated here to cause substrate inhibition of thiolase from *C. difficile* competing with acetoacetyl-CoA and this feature is shared by thiolases from different organisms **[21, 114, 115].** Indeed, it is obvious that thiolase was more sensitive towards the CoA-SH concentration than towards acetoacetyl-CoA. This indicates that thiolase activity depends on the presence of acetyl-CoA and the consumption of CoA-SH (the ratio of CoA-SH to acetyl-CoA). Thiolase from *C. acetobutylicum* is sensitive even at low ratios of CoA-SH to acetyl-CoA as well as to the accumulation of butyryl-CoA or ATP in high internal concentrations, suggesting that all these molecules may act as feedback inhibitors **[21]**. Additionally, the ratio of CoA-SH to acetyl-CoA has been shown to be an important regulator for NADH oxidation in *C. butyricum* and this is also thought to be an important mechanism of regulating the ratio of acetic to butyric acid **[158]**.

Other groups suggested that changes in the intracellular pH might be another important factor in regulating thiolase in *C. acetobutylicum*, since the activity of thiolase in *C. pasteurianum* falls with decreasing pH from an optimum of pH 8.0 **[23]**. ***Wiesenborn*** and co-workers confirmed that the pH profile in the condensation direction for *C. acetobutylicum* thiolase is very different from the one from *C. pasteurianum*, in that thiolase from *C. acetobutylicum* has high activity over a broad range of pH values from at least pH 5.5 to above 7.0 **[21]**. It has been shown, however, that in *C. acetobutylicum* batch fermentation the internal pH remains constant within the range of about pH 5.5 to 7.0 in the environment **[160]**.

Acetoacetyl-CoA is the product of the thiolase catalyzed reaction and provides a cross-over point in diverse metabolic pathways in microbes. It acts as a substrate for butanol/ butyrate and acetone formation, and is also a precursor in β-polyhydroxybutyrate (PHB) synthesis **[161, 162].** Thus, regulatory properties of BHBD might also be of relevance when the flux control of carbon through the cells is the matter of concern. BHBD from *C. difficile* was inhibited by acetoacetyl-CoA and the extent of inhibition was depended on the NADH concentration. It has been suggested that BHBD influences the ratio of butyrate to acetoacetate in acidogensis and the ratio of butanol to acetone during solventogenesis **[163-166]**. In the latter, decreased NADH and high concentration of acetoacetyl-CoA are known to redirect metabolic flux to acetoacetate/acetone pathway rather than converting it to β-hydroxybutyryl-CoA to complete butyrate/butanol pathway **[27, 117]**. It is known that the

enolate form of acetoacetyl-CoA acts as competitive inhibitor not only for BHBD but also for bovine liver crotonase **[29, 30]**. Acetocacetyl-CoA enolate structurally resembles both crotonyl-CoA and β-hydroxybutyryl-CoA and it can be considered a transition state structure for crotonase **[30]**. It is evident that acetoacetyl-CoA plays an important role in regulation of fatty acid pathways.

It is worth noting that in contrast to acetate, which is readily utilized at the end of aerobic growth **[78, 89, 90, 147]**, butyrate was not reoxidized at the end of the experiments. This is surprising, because butyrate can deliver more than twice the energy of acetate and both, activation as well as oxidation to acetyl-CoA might readily be achieved by the clostridial genes (Figure 5.4). Indeed, residual enzyme activities were detectable even in stationary cells. The capability of *E. coli* to utilize short-chain fatty acids has been extensively studied **[89, 90, 167]** and it turned out that the organisms inability to utilize short-chain fatty acids is restricted to C_4 and C_5. Longer and shorter fatty acids are readily metabolized. Indeed, it has been shown that not the metabolism of these compounds is impaired but the organism lacks uptake capability for these substrates **[111, 167-170]**.

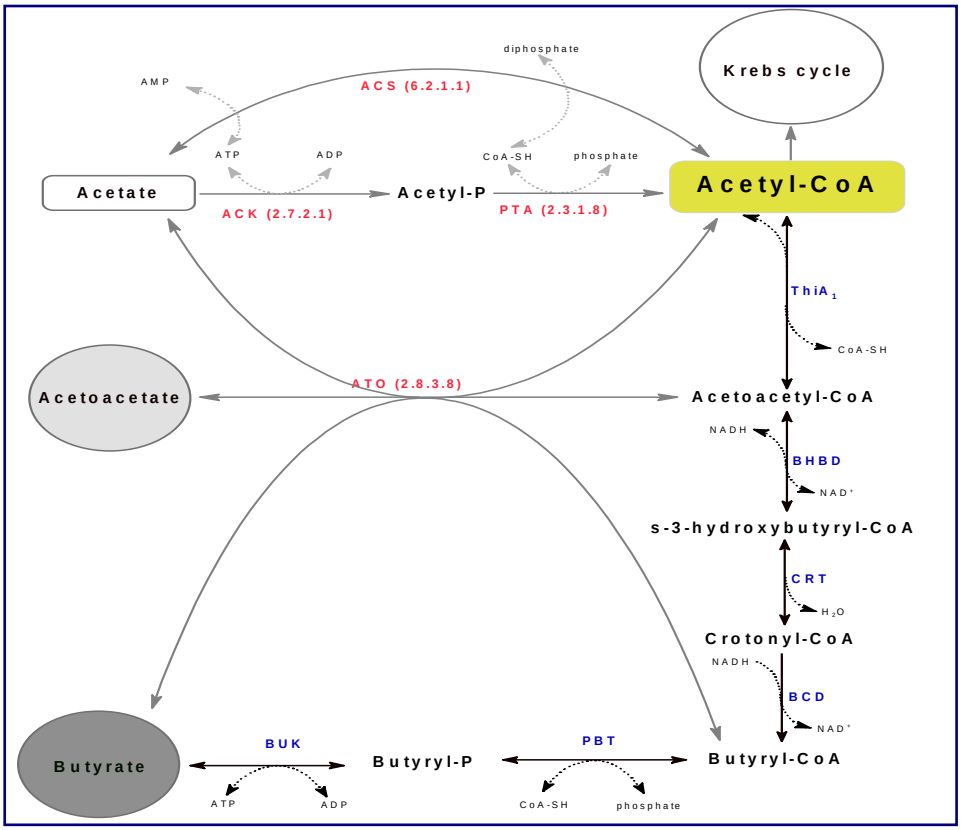


Fig.5.4: Proposed short-chain fatty acid uptake and utilization in modified *E. coli* strains. The blue enzyme symbols indicate the transferred butyrate pathway genes from clostridia, the red enzymes endogenous in *E. coli*.

The differences in short-chain fatty acid (acetate and butyrate) production patterns observed for the two modified strains as well as the control reflect the power of the novel analytic strategy applied in the current work. The cells harboring modular constructed, combinatorial plasmids can be readily generated and are valuable tools in order to study specific effects dedicated by the introduction of individual genes in a pre-defined metabolic pathway.

5.2.2. Rate-limiting step in butyrate fermentation

With the aim to figure out the rate-limiting step for butyrate production in modified *E. coli* strains, harboring butyrate biosynthetic pathway genes, enzyme activities were measured in cell free extracts obtained from butyrate fermentation biomass. Based on these data two rate-limiting steps in butyrate pathway fermentation can be identified. The first was the insufficient conversion of acetoacetyl-CoA to hydroxybutyryl-CoA with the rate-limiting enzyme BHBD. This observation agrees with ***Fischer** et al.*, who found that the intracellular metabolic flux in recombinant *E. coli* expressing clostridial butyrate and butanol biosynthetic genes is strongly affected by this enzyme **[66]**. The authors found that a shorting in conversion of acetoacetyl-CoA to hydroxybutyryl-CoA causes a metabolic bottle-neck and suggested that swapping NADH-requiring BHBD for NADPH-requiring 3-hydroxybutyryl-CoA reductase, such as PhaB might improve the situation **[66]**. The data presented herein for BHBD from *C. difficile* suggest that the problem is not associated with the expression level of the gene but caused by the regulatory properties of the enzyme. Although the crotonase gene is expressed at comparable low levels, the resulting enzyme is highly active and may cause concentrations of crotonyl-CoA to occur, which are inhibitory and must have detrimental effects on BHBD.

In line with this, the second rate-limiting step in the butyrate/butanol pathway is the conversion of crotonyl-CoA to butyryl-CoA that is catalyzed by butyryl-CoA dehydrogenase/ ETF complex. Several groups, who produced butanol from recombinant expression of clostridia genes, have reported that BCD/ETF complex is the rate-limiting enzyme, which is observed in much lower or not detectable levels at all as compared to other enzymes in the pathway **[56, 57, 66]**. These data rely on the individual measured enzyme activities **[56, 57]** as well as on the quantification of intracellular metabolites **[66]**. In addition, several groups noted oxygen sensitivity of butyryl-CoA dehydrogenases and the expression of BCD/ETF complexes from various origins under aerobic conditions led to low activities in *E. coli* due to

the proposed instability of the enzymes **[56, 57, 59, 66]**. Although the BCD complex of *C. difficile* is remarkable stable as compared to the *C. acetobutylicum* derived enzyme, its activity in cell lysates of recombinant strains was still the lowest of the pathway. It should be noted, however, that the failure to measure BCD from *C. acetobutylicum* in the crotonyl-CoA reducing direction enforced the measurement of oxidative reaction for comparison and that this values not at all relates to the physiological reaction of interest.

Since the availability of NADH increases **[171, 172]**, one might expect higher butyrate yields from glucose under anoxic conditions. It should be noted, however, that growth experiments using *E. coli* BUT_D aiming anoxic expression of clostridial genes actually caused lower concentrations of butyrate to accumulate in the medium than in aerobically grown cultures. This observation is detrimental to the proposal that product formation is solely driven by the availability of NADH for substrate reduction. In fact it is known that the ratio of NADH to NAD^+ rises in *E coli* from 0.05 to above 0.75 upon transfer from an oxic to an anoxic environment **[173]**, a fact that might enhance butyrate formation. Thus, it seems reasonable to conclude that the unique properties of bi-forking BCD could be responsible for this observation. The *in vitro* assays revealed a strict requirement of BCD for oxidized ferredoxin in anoxic tests, while under air also dioxygen was effective as a co-substrate for crotonyl-CoA reduction. Under anoxic conditions, the reoxidation of ferredoxin is predicted to be severely impaired in pure cultures. While endogenous ferredoxin:NADP oxidoreductase can readily provide oxidized ferredoxin or oxygen itself may act as co-oxidant for BCD *in vivo*, under anoxic conditions the intracellular redox potential is much lower and the reoxidation of ferredoxin becomes a more difficult task. In fact, reoxidation of ferredoxin under anoxic conditions is almost entirely achieved by hydrogenases in *E. coli* **[174-176]**, which can, however, act in pure cultures only within narrow thermodynamic limits.

5.3. Implications of the novel findings for understanding of clostridial butyrate metabolism

Five strains of *C. difficile* were previously found to be capable of growing either chemo-lithotrophically with H_2 and CO_2 or chemo-organotrophoically with glucose, fructose, maltose, mannose and cellobiose **[177]**. Based on these findings, *C. difficile* was classified as an acetogenic bacterium. The intermediate-key between those two metabolisms is the energy conservation via reduced ferredoxin. The novel findings in this study combined with the genome-derived data, a new hypothesis regarding the energy conservation in acetogenesis by *C. difficile* can be formulated. Under autotrophic growth conditions, the first step is to

generate reduced ferredoxin and this process is provided by hydrogenases utilizing H_2 as electron donor. Reduced ferredoxin thus formed is then used for two proposals: Firstly, CO_2 is reduced via Carbon-monoxide dehydrogenase/ acetyl-CoA synthase (EC 1.2.99.2) to provide CO in the Wood-Ljungdahl pathway. Secondly, the reduced ferredoxin is shuttled by a membrane-bound protein (RNF-complex) to produce NADH that is needed in the Eastern branch of acetogenesis **[178, 179]**. The reoxidation of ferredoxin via RNF-complex can conserve additional energy via electron transport phosphorylation to synthesize ATP, in which the released electrons are used to create a proton motive force across the cytoplasma membrane **[17, 179, 180]**.

Under heterotrophic growth conditions, *C. difficile* is capable of growing in a defined medium with glucose as the sole energy and carbon source **[181]**. Butyrate formation in clostridia was accepted for decades to provide an electron sink for glucose oxidation. During glucose oxidation, pyruvate is oxidized by pyruvate: ferredoxin oxidoreductase (PFOR, EC 1.2.7.1) to provide reduced ferredoxin and CO_2, which are essential in heterotrophic growth to provide an additional molecule of acetyl-CoA **[179, 182, 183]**. Consequently, up to 3 mole of acetyl-CoA are formed per mol glucose **[179]**. In the reductive branch of butyrate pathway, reduction of crotonyl-CoA to butyryl-CoA is mediated by a bi-forking butyryl-CoA dehydrogenase/ ETF complex, which couples the exergonic reduction of crotonyl-CoA with the endergonic reduction of ferredoxin and provide additional energy conservation by reoxidation of reduced ferredoxin via RNF- complex **[17, 33, 179]**.

The anoxic oxidation of short-chain fatty acids in clostridia may also reflect the central role of the RNF-complex. Anoxic butyrate degradation in *C. difficile* yields two molecules of acetyl-CoA that allows substrate level phosphorylation of ADP within acetyl-phosphate transferase and acetate kinase. The limiting step in this pathway is the oxidation of butyryl-CoA to crotonyl-CoA with NAD^+. This step is endergonic, while all other steps might proceed smoothly. This endergonic reaction can only be driven by coupling with exergonic oxidation of reduced ferredoxin, which is provided by H_2 oxidation and RNF-complex. Acetogenesis serves as electron sink necessary to balance redox potential and the strict metabolic coupling of butyrate oxidation and acetogenesis has been convincingly shown in thermophilic mixed cultures when methanogenesis was suppressed **[184]**

5.4. Outlook

The two butyrate modules, *pASG-BUT*$_D$ and *pASG-BUT*$_{AC}$ are readily modified replacing *pbt* and *buk* genes, by *adhe* (Fig. 5.6). The *adhe* encoding a bifunctional aldehyde/alcohol dehydrogenase is then able to convert butyryl-CoA to butan-1-ol consuming two additional NADH (Fig.5.5) and enables conversion of acetyl-CoA into this commercially interesting solvent. The gene from *C. acetobutylicum* (accession number CA_P0035) was cloned and sequenced, and its integration into the vector is presently on the way.

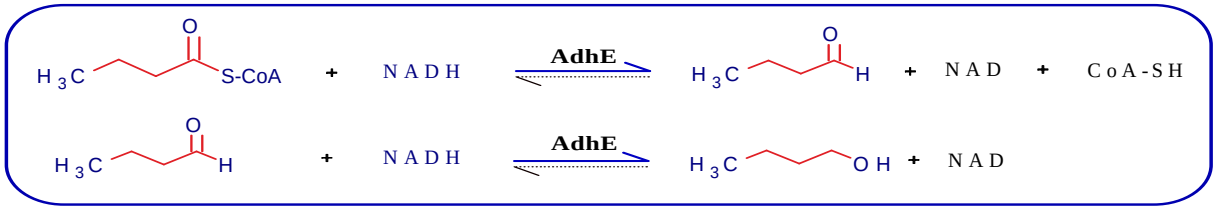


Fig.5.5: The enzymatic reaction of the bi-functional enzymes ADHE.

Given that these novel modules confer butan-1-ol production capability to *E. coli*, the solvent producing capability can be coupled to alternative feedstocks in combinatorial approaches. Thereby, in a truly synthetic biological approach, dedicated strains for the specific conversion of various constituents of biomass will be generated, which are not found naturally in known organisms and might be highly interesting in technical applications aiming conversion of these renewable sources in solvents. Among the most interesting targets for nutrients, which are available in large quantities and not utilized for alternative use (e.g. human or animal nutrition) are lignocelluloses, which are available in rice straw in huge quantities all around the world, or raw glycerol, which is a low value by-product of biodiesel production from plant oils.

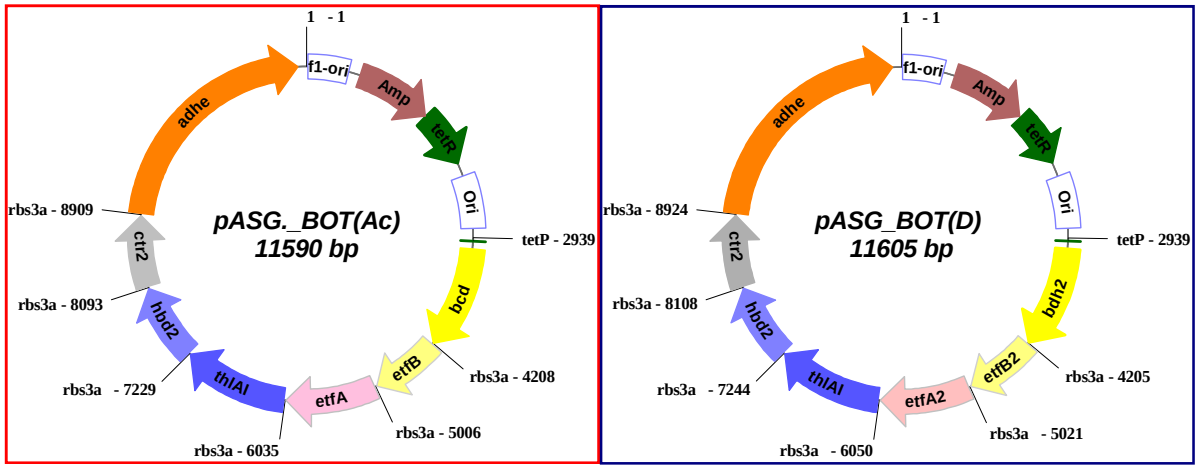


Fig.5.6: Expression plasmids for butan-1-ol pathway modules.

6. References

1. **The.** Free Online Dictionary, http://www.thefreedictionary.com/Clostridium+difficile.
2. **Hull, M. W., and P. L. Beck.** (2004). *Clostridium difficile*-associated colitis. Can Fam Physician **50:**1536-40, 1543-5.
3. **Hookman, P., and J. S. Barkin.** (2009). *Clostridium difficile* associated infection, diarrhea and colitis. World J Gastroenterol **15:**1554-80.
4. **Yassin, S. F., T. M. Young-Fadok, N. N. Zein, and D. S. Pardi.** (2001). *Clostridium difficile*-associated diarrhea and colitis. Mayo Clin Proc **76:**725-30.
5. **Gerding, D. N., S. Johnson, L. R. Peterson, M. E. Mulligan, and J. Silva, Jr.** (1995). *Clostridium difficile*-associated diarrhea and colitis. Infect Control Hosp Epidemiol **16:**459-77.
6. **Bartlett, J. G.** (1994). *Clostridium difficile*: history of its role as an enteric pathogen and the current state of knowledge about the organism. Clin Infect Dis **18 Suppl 4:**S265-72.
7. **Kelly, C. P., C. Pothoulakis, and J. T. LaMont.** (1994). *Clostridium difficile* colitis. N Engl J Med **330:**257-62.
8. **Warny, M., A. C. Keates, S. Keates, I. Castagliuolo, J. K. Zacks, S. Aboudola, A. Qamar, C. Pothoulakis, J. T. LaMont, and C. P. Kelly.** (2000). p38 MAP kinase activation by *Clostridium difficile* toxin A mediates monocyte necrosis, IL-8 production, and enteritis. J Clin Invest **105:**1147-56.
9. **Paredes, C. J., K. V. Alsaker, and E. T. Papoutsakis.** (2005). A comparative genomic view of clostridial sporulation and physiology. Nat Rev Microbiol **3:**969-78.
10. **Sebaihia, M., B. W. Wren, P. Mullany, N. F. Fairweather, N. Minton, R. Stabler, N. R. Thomson, A. P. Roberts, A. M. Cerdeno-Tarraga, H. Wang, M. T. Holden, A. Wright, C. Churcher, M. A. Quail, S. Baker, N. Bason, K. Brooks, T. Chillingworth, A. Cronin, P. Davis, L. Dowd, A. Fraser, T. Feltwell, Z. Hance, S. Holroyd, K. Jagels, S. Moule, K. Mungall, C. Price, E. Rabbinowitsch, S. Sharp, M. Simmonds, K. Stevens, L. Unwin, S. Whithead, B. Dupuy, G. Dougan, B. Barrell, and J. Parkhill.** (2006). The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. Nat Genet **38:**779-86.
11. **Sebaihia M., S. L., Wellcome Trust Sanger Institute,, and H. Wellcome Trust Genome Campus, Cambridge, CB10 1SA, UNITED KINGDOM.** (2005). *Clostridium difficile* 630 complete genome. http://www.ebi.ac.uk/Tools/dbfetch/dbfetch?db=embl&id=AM180355&style=raw.

12. **Nolling, J., G. Breton, M. V. Omelchenko, K. S. Makarova, Q. Zeng, R. Gibson, H. M. Lee, J. Dubois, D. Qiu, J. Hitti, Y. I. Wolf, R. L. Tatusov, F. Sabathe, L. Doucette-Stamm, P. Soucaille, M. J. Daly, G. N. Bennett, E. V. Koonin, and D. R. Smith.** (2001). Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. J Bacteriol **183:**4823-38.
13. **Sebaihia, M., M. W. Peck, N. P. Minton, N. R. Thomson, M. T. Holden, W. J. Mitchell, A. T. Carter, S. D. Bentley, D. R. Mason, L. Crossman, C. J. Paul, A. Ivens, M. H. Wells-Bennik, I. J. Davis, A. M. Cerdeno-Tarraga, C. Churcher, M. A. Quail, T. Chillingworth, T. Feltwell, A. Fraser, I. Goodhead, Z. Hance, K. Jagels, N. Larke, M. Maddison, S. Moule, K. Mungall, H. Norbertczak, E. Rabbinowitsch, M. Sanders, M. Simmonds, B. White, S. Whithead, and J. Parkhill.** (2007). Genome sequence of a proteolytic (Group I) *Clostridium botulinum* strain Hall A and comparative analysis of the clostridial genomes. Genome Res **17:**1082-92.
14. **Shimizu, T., K. Ohtani, H. Hirakawa, K. Ohshima, A. Yamashita, T. Shiba, N. Ogasawara, M. Hattori, S. Kuhara, and H. Hayashi.** (2002). Complete genome sequence of *Clostridium perfringens*, an anaerobic flesh-eater. Proc Natl Acad Sci U S A **99:**996-1001.
15. **Bruggemann, H., S. Baumer, W. F. Fricke, A. Wiezer, H. Liesegang, I. Decker, C. Herzberg, R. Martinez-Arias, R. Merkl, A. Henne, and G. Gottschalk.** (2003). The genome sequence of *Clostridium tetani*, the causative agent of tetanus disease. Proc Natl Acad Sci U S A **100:**1316-21.
16. **Jain, S., R. L. Graham, G. McMullan, and N. G. Ternan.** (2010). Proteomic analysis of the insoluble subproteome of *Clostridium difficile* strain 630. FEMS Microbiol Lett **312:**151-9.
17. **Herrmann, G., E. Jayamani, G. Mai, and W. Buckel.** (2008). Energy conservation via electron-transferring flavoprotein in anaerobic bacteria. J Bacteriol **190:**784-91.
18. **Lynen, F., and S. Ochoa.** (1953). Enzymes of fatty acid metabolism. Biochim Biophys Acta **12:**299-314.
19. **Stim-Herndon, K. P., D. J. Petersen, and G. N. Bennett.** (1995). Characterization of an acetyl-CoA C-acetyltransferase (thiolase) gene from *Clostridium acetobutylicum* ATCC 824. Gene **154:**81-5.
20. **Petersen, D. J., and G. N. Bennett.** (1991). Cloning of the *Clostridium acetobutylicum* ATCC 824 acetyl coenzyme A acetyltransferase (thiolase; EC 2.3.1.9) gene. Appl Environ Microbiol **57:**2735-41.

21. **Wiesenborn, D. P., F. B. Rudolph, and E. T. Papoutsakis.** (1988). Thiolase from *Clostridium acetobutylicum* ATCC 824 and Its Role in the Synthesis of Acids and Solvents. Appl Environ Microbiol **54:**2717-22.
22. **Sliwkowski, M. X., and M. G. Hartmanis.** (1984). Simultaneous single-step purification of thiolase and NADP-dependent 3-hydroxybutyryl-CoA dehydrogenase from Clostridium kluyveri. Anal Biochem **141:**344-7.
23. **Berndt, H., and H. G. Schlegel.** (1975). Kinetics and properties of beta-ketothiolase from *Clostridium pasteurianum*. Arch Microbiol **103:**21-30.
24. **Madan, V. K., P. Hillmer, and G. Gottschalk.** (1973). Purification and properties of NADP-dependent L(+)-3-hydroxybutyryl-CoA dehydrogenase from *Clostridium kluyveri*. Eur J Biochem **32:**51-6.
25. **Boynton, Z. L., G. N. Bennet, and F. B. Rudolph.** (1996). Cloning, sequencing, and expression of clustered genes encoding beta-hydroxybutyryl-coenzyme A (CoA) dehydrogenase, crotonase, and butyryl-CoA dehydrogenase from *Clostridium acetobutylicum* ATCC 824. J Bacteriol **178:**3015-24.
26. **Mullany, P., C. L. Clayton, M. J. Pallen, R. Slone, A. al-Saleh, and S. Tabaqchali.** (1994). Genes encoding homologues of three consecutive enzymes in the butyrate/butanol-producing pathway of *Clostridium acetobutylicum* are clustered on the *Clostridium difficile* chromosome. FEMS Microbiol Lett **124:**61-7.
27. **Hartmanis, M. G., and S. Gatenbeck.** (1984). Intermediary Metabolism in Clostridium acetobutylicum: Levels of Enzymes Involved in the Formation of Acetate and Butyrate. Appl Environ Microbiol **47:**1277-83.
28. **Engel, C. K., T. R. Kiema, J. K. Hiltunen, and R. K. Wierenga.** (1998). The crystal structure of enoyl-CoA hydratase complexed with octanoyl-CoA reveals the structural adaptations required for binding of a long chain fatty acid-CoA molecule. J Mol Biol **275:**847-59.
29. **Engel, C. K., M. Mathieu, J. P. Zeelen, J. K. Hiltunen, and R. K. Wierenga.** (1996). Crystal structure of enoyl-coenzyme A (CoA) hydratase at 2.5 angstroms resolution: A spiral fold defines the CoA-binding pocket. EMBO J **15:**5135-45.
30. **Waterson, R. M., and R. L. Hill.** (1972). Enoyl coenzyme A hydratase (crotonase). Catalytic properties of crotonase and its possible regulatory role in fatty acid oxidation. J Biol Chem **247:**5258-65.

31. **Waterson, R. M., G. M. Hass, and R. L. Hill.** (1972). Enoyl coenzyme A hydratase (crotonase). Enhancement of the rate of hydration of crotonylpantetheine by coenzyme A and related compounds. J Biol Chem **247:**5252-7.
32. **Waterson, R. M., F. J. Castellino, G. M. Hass, and R. L. Hill.** (1972). Purification and Characterization of Crotonase from *Clostridium acetobutylicum*. J Biol Chem **247:**5266-71.
33. **Li, F., J. Hinderberger, H. Seedorf, J. Zhang, W. Buckel, and R. K. Thauer.** (2008). Coupled ferredoxin and crotonyl coenzyme A (CoA) reduction with NADH catalyzed by the butyryl-CoA dehydrogenase/Etf complex from *Clostridium kluyveri*. J Bacteriol **190:**843-50.
34. **Djordjevic, S., C. P. Pace, M. T. Stankovich, and J. J. Kim.** (1995). Three-dimensional structure of butyryl-CoA dehydrogenase from *Megasphaera elsdenii*. Biochemistry **34:**2163-71.
35. **Van Berkel, W. J., W. A. Van den Berg, and F. Muller.** (1988). Large-scale preparation and reconstitution of apo-flavoproteins with special reference to butyryl-CoA dehydrogenase from *Megasphaera elsdenii*. Hydrophobic-interaction chromatography. Eur J Biochem **178:**197-207.
36. **Whitfield, C. D., and S. G. Mayhew.** (1974). Purification and properties of electron-transferring flavoprotein from *Peptostreptococcus elsdenii*. J Biol Chem **249:**2801-10.
37. **Engel, P. C., and V. Massey.** (1971). Green butyryl-coenzyme A dehydrogenase. An enzyme-acyl-coenzyme A complex. Biochem J **125:**889-902.
38. **Oultram, J. D., I. D. Burr, M. J. Elmore, and N. P. Minton.** (1993). Cloning and sequence analysis of the genes encoding phosphotransbutyrylase and butyrate kinase from *Clostridium acetobutylicum* NCIMB 8052. Gene **131:**107-12.
39. **Thompson, D. K., and J. S. Chen.** (1990). Purification and properties of an acetoacetyl coenzyme A-reacting phosphotransbutyrylase from *Clostridium beijerinckii* ("*Clostridium butylicum*") NRRL B593. Appl Environ Microbiol **56:**607-13.
40. **Wiesenborn, D. P., F. B. Rudolph, and E. T. Papoutsakis.** (1989). Phosphotransbutyrylase from *Clostridium acetobutylicum* ATCC 824 and its role in acidogenesis. Appl Environ Microbiol **55:**317-22.
41. **Cary, J. W., D. J. Petersen, E. T. Papoutsakis, and G. N. Bennett.** (1988). Cloning and expression of *Clostridium acetobutylicum* phosphotransbutyrylase and butyrate kinase genes in *Escherichia coli*. J Bacteriol **170:**4613-8.

42. **Valentine, R. C., and R. S. Wolfe.** (1960). Purification and role of phosphotransbutyrylase. J Biol Chem **235:**1948-52.
43. **Huang, K. X., S. Huang, F. B. Rudolph, and G. N. Bennett.** (2000). Identification and characterization of a second butyrate kinase from *Clostridium acetobutylicum* ATCC 824. J Mol Microbiol Biotechnol **2:**33-8.
44. **Oultram, J. D., I. D. Burr, M. J. Elmore, and N. P. Minton.** (1993). Cloning and sequence analysis of the genes encoding phoshotransbutyrylase and butyrate kinase from *Clostridium acetobutylicum* NCIMB 8052. Gene **131:**07-112.
45. **Hartmanis, M. G. N.** (1987). Butyrate Kinase from *Clostridium acetobutylicum*. J Biol Chem. **262:**617-621.
46. **Twarog, R., and R. S. Wolfe.** (1963). Role of butyryl phosphate in the energy metabolism of *Clostridium tetanomorphum*. J. Bacteriol. **86:**112-117.
47. **Twarog, R., and R. S. Wolfe.** (1962). Enzymatic Phosphorylation of Butyrate. J Biol Chem. **237:**2474- 2477.
48. **Lengeler, J., Drews, G., and Schlegel, H. .** (1999). Biology of the Prokaryotes. Blackwell Publishing. UK.
49. **Boiangiu, C. D., E. Jayamani, D. Brugel, G. Herrmann, J. Kim, L. Forzi, R. Hedderich, I. Vgenopoulou, A. J. Pierik, J. Steuber, and W. Buckel.** (2005). Sodium ion pumps and hydrogen production in glutamate fermenting anaerobic bacteria. J Mol Microbiol Biotechnol **10:**105-19.
50. **Seedorf, H., W. F. Fricke, B. Veith, H. Bruggemann, H. Liesegang, A. Strittmatter, M. Miethke, W. Buckel, J. Hinderberger, F. Li, C. Hagemeier, R. K. Thauer, and G. Gottschalk.** (2008). The genome of *Clostridium kluyveri*, a strict anaerobe with unique metabolic features. Proc Natl Acad Sci U S A **105:**2128-33.
51. **Escherich, T.** (1885). Die darmbakterien des neugeborenen und sauglings. Fortshr. Med. 3:5-15-522, 547-554.
52. **NCBI., T. B.**, posting date. Escherichia. [Online.]
53. **Vogt, R. L., and L. Dippold.** (2005). *Escherichia coli* O157:H7 outbreak associated with consumption of ground beef, June-July 2002. Public Health Rep **120:**174-8.
54. **Choi, J. H., Keum, K.C. and Lee, S.Y.** (2006). Production of recombinant proteins by high cell density culture of *Escherichia coli*. Chemical Engineering Science **61:**876 – 885.

55. **Yoon, S. H., S. K. Kim, and J. F. Kim.** (2010). Secretory production of recombinant proteins in *Escherichia coli*. Recent Pat Biotechnol **4:**23-9.
56. **Nielsen, D. R., E. Leonard, S. H. Yoon, H. C. Tseng, C. Yuan, and K. L. Prather.** (2009). Engineering alternative butanol production platforms in heterologous bacteria. Metab Eng **11:**262-73.
57. **Inui, M., M. Suda, S. Kimura, K. Yasuda, H. Suzuki, H. Toda, S. Yamamoto, S. Okino, N. Suzuki, and H. Yukawa.** (2008). Expression of *Clostridium acetobutylicum* butanol synthetic genes in *Escherichia coli*. Appl Microbiol Biotechnol **77:**1305-16.
58. **Atsumi, S., and J. C. Liao.** (2008). Metabolic engineering for advanced biofuels production from *Escherichia coli*. Curr Opin Biotechnol **19:**414-9.
59. **Atsumi, S., A. F. Cann, M. R. Connor, C. R. Shen, K. M. Smith, M. P. Brynildsen, K. J. Chou, T. Hanai, and J. C. Liao.** (2008). Metabolic engineering of *Escherichia coli* for 1-butanol production. Metab Eng **10:**305-11.
60. **Tseng, H. C., C. H. Martin, D. R. Nielsen, and K. L. Prather.** (2009). Metabolic engineering of *Escherichia coli* for enhanced production of (R)- and (S)-3-hydroxybutyrate. Appl Environ Microbiol **75:**3137-45.
61. **Do, P. M., A. Angerhofer, I. Hrdy, L. Bardonova, L. O. Ingram, and K. T. Shanmugam.** (2009). Engineering *Escherichia coli* for fermentative dihydrogen production: Potential role of NADH-ferredoxin oxidoreductase from the hydrogenosome of anaerobic protozoa. Appl Biochem Biotechnol **153:**21-33.
62. **Jojima, T., M. Inui, and H. Yukawa.** (2008). Production of isopropanol by metabolically engineered *Escherichia coli*. Appl Microbiol Biotechnol **77:**1219-24.
63. **Chavez-Bejar, M. I., A. R. Lara, H. Lopez, G. Hernandez-Chavez, A. Martinez, O. T. Ramirez, F. Bolivar, and G. Gosset.** (2008). Metabolic engineering of *Escherichia coli* for L-tyrosine production by expression of genes coding for the chorismate mutase domain of the native chorismate mutase-prephenate dehydratase and a cyclohexadienyl dehydrogenase from *Zymomonas mobilis*. Appl Environ Microbiol **74:**3284-90.
64. **Hanai, T., S. Atsumi, and J. C. Liao.** (2007). Engineered synthetic pathway for isopropanol production in *Escherichia coli*. Appl Environ Microbiol **73:**7814-8.
65. **Lee, S. Y.** (1996). High cell-density culture of *Escherichia coli*. Trends in Biotechnology **14:**98-105.
66. **Fischer, C. R., H. C. Tseng, M. Tai, K. L. Prather, and G. Stephanopoulos.** (2010). Assessment of heterologous butyrate and butanol pathway activity by measurement of

intracellular pathway intermediates in recombinant *Escherichia coli*. Appl Microbiol Biotechnol **88:**265-75.

67. **Na, D., T. Y. Kim, and S. Y. Lee.** (2010). Construction and optimization of synthetic pathways in metabolic engineering. Curr Opin Microbiol **13:**363-70.
68. **Clomburg, J. M., and R. Gonzalez.** (2010). Biofuel production in *Escherichia coli*: The role of metabolic engineering and synthetic biology. Appl Microbiol Biotechnol **86:**419-34.
69. **Atsumi, S., and J. C. Liao.** (2008). Directed evolution of *Methanococcus jannaschii* citramalate synthase for biosynthesis of 1-propanol and 1-butanol by *Escherichia coli*. Appl Environ Microbiol **74:**7802-8.
70. **Yomano, L. P., S. W. York, S. Zhou, K. T. Shanmugam, and L. O. Ingram.** (2008). Re-engineering *Escherichia coli* for ethanol production. Biotechnol Lett **30:**2097-103.
71. **Withers, S. T., S. S. Gottlieb, B. Lieu, J. D. Newman, and J. D. Keasling.** (2007). Identification of isopentenol biosynthetic genes from *Bacillus subtilis* by a screening method based on isoprenoid precursor toxicity. Appl Environ Microbiol **73:**6277-83.
72. **Andreessen, B., A. B. Lange, H. Robenek, and A. Steinbuchel.** (2010). Conversion of glycerol to poly(3-hydroxypropionate) in recombinant *Escherichia coli*. Appl Environ Microbiol **76:**622-6.
73. **Shen, C. R., and J. C. Liao.** (2008). Metabolic engineering of *Escherichia coli* for 1-butanol and 1-propanol production via the keto-acid pathways. Metab Eng **10:**312-20.
74. **Lemuth, K., T. Hardiman, S. Winter, D. Pfeiffer, M. A. Keller, S. Lange, M. Reuss, R. D. Schmid, and M. Siemann-Herzberg.** (2008). Global transcription and metabolic flux analysis of *Escherichia coli* in glucose-limited fed-batch cultivations. Appl Environ Microbiol **74:**7002-15.
75. **Aon, J. C., R. J. Caimi, A. H. Taylor, Q. Lu, F. Oluboyede, J. Dally, M. D. Kessler, J. J. Kerrigan, T. S. Lewis, L. A. Wysocki, and P. S. Patel.** (2008). Suppressing posttranslational gluconoylation of heterologous proteins by metabolic engineering of *Escherichia coli*. Appl Environ Microbiol **74:**950-8.
76. **Bernal, V., B. Masdemont, P. Arense, M. Canovas, and J. L. Iborra.** (2007). Redirecting metabolic fluxes through cofactor engineering: Role of CoA-esters pool during L(-)-carnitine production by *Escherichia coli*. J Biotechnol **132:**110-7.
77. **Lee, S. J., D. Y. Lee, T. Y. Kim, B. H. Kim, J. Lee, and S. Y. Lee.** (2005). Metabolic engineering of *Escherichia coli* for enhanced production of succinic acid, based on

genome comparison and in silico gene knockout simulation. Appl Environ Microbiol **71:**7880-7.

78. **Holms, H.** (1996). Flux analysis and control of the central metabolic pathways in *Escherichia coli*. FEMS Microbiol Rev **19:**85-116.
79. **Rutherford, B. J., R. H. Dahl, R. E. Price, H. L. Szmidt, P. I. Benke, A. Mukhopadhyay, and J. D. Keasling.** (2010). Functional genomic study of exogenous n-butanol stress in *Escherichia coli*. Appl Environ Microbiol **76:**1935-45.
80. **Sorensen, H. P., and K. K. Mortensen.** (2005). Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. J Biotechnol **115:**113-28.
81. **di Salvo, M. L., S. Hunt, and V. Schirch.** (2004). Expression, purification, and kinetic constants for human and *Escherichia coli* pyridoxal kinases. Protein Expr Purif **36:**300-6.
82. **DeLisa, M. P., J. Li, G. Rao, W. A. Weigand, and W. E. Bentley.** (1999). Monitoring GFP-operon fusion protein expression during high cell density cultivation of *Escherichia coli* using an on-line optical sensor. Biotechnol Bioeng **65:**54-64.
83. **Arthur, P. K., L. J. Alvarado, and T. K. Dayie.** (2011). Expression, purification and analysis of the activity of enzymes from the pentose phosphate pathway. Protein Expr Purif **76(2):229-37:**9.
84. **Teng, C., H. Jia, Q. Yan, P. Zhou, and Z. Jiang.** (2011). High-level expression of extracellular secretion of a beta-xylosidase gene from *Paecilomyces thermophila* in *Escherichia coli*. Bioresour Technol **102:**1822-30.
85. **Helianti, I., N. Nurhayati, M. Ulfah, B. Wahyuntari, and S. Setyahadi.** (2010). Constitutive high level expression of an endoxylanase gene from the newly isolated *Bacillus subtilis* AQ1 in *Escherichia coli*. J Biomed Biotechnol **2010**.
86. **Hsieh, L. S., C. S. Yeh, H. C. Pan, C. Y. Cheng, C. C. Yang, and P. D. Lee.** (2010). Cloning and expression of a phenylalanine ammonia-lyase gene (BoPAL2) from *Bambusa oldhamii* in *Escherichia coli* and *Pichia pastoris*. Protein Expr Purif **71:**224-30.
87. **Shiloach, J., S. Reshamwala, S. B. Noronha, and A. Negrete.** (2010). Analyzing metabolic variations in different bacterial strains, historical perspectives and current trends--example *E. coli*. Curr Opin Biotechnol **21:**21-6.
88. **Lequeux, G., J. Beauprez, J. Maertens, E. Van Horen, W. Soetaert, E. Vandamme, and P. A. Vanrolleghem.** (2010). Dynamic metabolic flux analysis demonstrated on cultures where the limiting substrate is changed from carbon to nitrogen and vice versa. J Biomed Biotechnol **2010**.

89. **Salanitro, J. P., and W. S. Wegener.** (1971). Growth of *Escherichia coli* on short-chain fatty acids: Growth characteristics of mutants. J Bacteriol **108:**885-92.
90. **Salanitro, J. P., and W. S. Wegener.** (1971). Growth of *Escherichia coli* on short-chain fatty acids: nature of the uptake system. J Bacteriol **108:**893-901.
91. **Weeks, G., M. Shapiro, R. O. Burns, and S. J. Wakil.** (1969). Control of fatty acid metabolism. I. Induction of the enzymes of fatty acid oxidation in *Escherichia coli*. J Bacteriol **97:**827-36.
92. **Cho, B. K., E. M. Knight, and B. O. Palsson.** (2006). Transcriptional regulation of the fad regulon genes of *Escherichia coli* by ArcA. Microbiology **152:**2207-19.
93. **Uede, Y.** Jun.28, 1988 (1988). Process for preparing solid acetyl phosphate salt.
94. **Simon E. J ., and D. Hemis.** (1953). The Preparation of S-Succinyl Coenzyme A'. J. Am. Chem. Soc **75:**2520.
95. **Selmer, T., and W. Buckel.** (1999). Oxygen Exchange between Acetate and the Catalytic Glutamate Residue in Glutaconate CoA-transferase from *Acidaminococcus fermentans*. IMPLICATIONS FOR THE MECHANISM OF CoA-ESTER HYDROLYSIS. Journal of Biological Chemistry **274:**20772 - 20778.
96. **Inoue, H., H. Nojima, and H. Okayama.** (1990). High efficiency transformation of *Escherichia coli* with plasmids. Gene **96:**23-8.
97. **Sanger, F., S. Nicklen, and A. R. Coulson.** (1977). DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A **74:**5463-7.
98. **Bradford, M. M.** (1976). A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. Analytical Biochemistry **72:**248- 254.
99. **Gill, S. C., and P. H. von Hippel.** (1989). Calculation of protein extinction coefficients from amino acid sequence data. Anal Biochem **182:**319-26.
100. **Laemmli, U. K.** (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. . Nature (London) **227:**680-685.
101. **Lehman, T. C., D. E. Hale, A. Bhala, and C. Thorpe.** (1990). An acyl-coenzyme A dehydrogenase assay utilizing the ferricenium ion. Anal Biochem **186:**280-4.
102. **Storer, A. C., and A. Cornish-Bowden.** (1974). The kinetics of coupled enzyme reactions. Applications to the assay of glucokinase, with glucose 6-phosphate dehydrogenase as coupling enzyme. Biochem J **141:**205-9.
103. **Kunst A., Draeger B., and Z. J.** (1984). In Methods of Enzymatic Analysis, 3 ed, vol. 6. Bergmeyer, H.U., ed, VCH, Weinheim-Deerfield Beach, FL-Basel.

104. **Schiffels, J., M. E. Baumann, and T. Selmer.** (2011). Facile analysis of short-chain fatty acids as 4-nitrophenyl esters in complex anaerobic fermentation samples by high performance liquid chromatography. J Chromatogr A. (in press)

105. **Heijnen, J. J.** (2010). Mass Balances, Rates, and Experiments *In* C. D. Smolke (ed.), The Metabolic Pathway Engineering Handbook: Fundamentals, 1st ed. CRC Press Taylor & Francis Group, Boca Raton, New York.

106. **Sharp, P. M., and W. H. Li.** (1987). The codon Adaptation Index a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res **15:**1281-95.

107. **Aboulnaga, E., O. Pinkenburg, J. Schiffels, R. Fischer, A. El-Refai, W. Buckel and T. Selmer.** (2011). *Clostridium difficile* takes advantage of butyryl-CoA dehydrogenase driven electron transport phosphorylation. In preparation.

108. **Shin, R., M. Suzuki, and Y. Morishita.** (2002). Influence of intestinal anaerobes and organic acids on the growth of enterohaemorrhagic *Escherichia coli* O157:H7. J Med Microbiol **51:**201-6.

109. **Schiffels, J., M. E. M. Baumann,T. Selmer** (2011). Facile analysis of short-chain fatty acids as 4-nitrophenyl esters in complex anaerobic fermentation samples by high performance liquid chromatography. (Under revision).

110. **Duncombe, G. R., and F. E. Frerman.** (1976). Molecular and catalytic properties of the acetoacetyl-coenzyme A thiolase of *Escherichia coli*. Arch Biochem Biophys **176:**159-70.

111. **Feigenbaum, J., and H. Schulz.** (1975). Thiolases of *Escherichia coli*: purification and chain length specificities. J Bacteriol **122:**407-11.

112. **Hartmanis, M. G., and T. C. Stadtman.** (1982). Isolation of a selenium-containing thiolase from *Clostridium kluyveri*: identification of the selenium moiety as selenomethionine. Proc Natl Acad Sci U S A **79:**4912-6.

113. **Reddick, J. J., and J. K. Williams.** (2008). The mmgA gene from *Bacillus subtilis* encodes a degradative acetoacetyl-CoA thiolase. Biotechnol Lett **30:**1045-50.

114. **Middleton, B.** (1974). The kinetic mechanism and properties of the cytoplasmic acetoacetyl-coenzyme A thiolase from rat liver. Biochem J **139:**109-21.

115. **Nishimura, T., T. Saito, and K. Tomita.** (1978). Purification and properties of beta-ketothiolase from *Zoogloea ramigera*. Arch Microbiol **116:**21-7.

116. **Suzuki, F., W. L. Zahler, and D. W. Emerich.** (1987). Acetoacetyl-CoA thiolase of *Bradyrhizobium japonicum* bacteroids: Purification and properties. Arch Biochem Biophys **254:**272-81.

117. **Salleh, M. M., L. S. Tsuey, and A. B. Ariff.** (2008). The Profile of Enzymes Relevant to Solvent Production during Direct Fermentation of Sago Starch by *Clostridium saccharobutylicum* P262 Utilizing Different pH Control Strategies. Biotechnology and Bioprocess Engineering **13:**33-39.

118. **Lee, C., H. Goerisch, and R. Zocher.** (2000). The Kinetic Investigation of D-Hydroxyisovalerate Dehydrogenase from *Fusarium sambucinum*. Journal of Biochemistry and Molecular Biology **33:**228-233.

119. **Shoemark, D. K., M. J. Cliff, R. B. Sessions, and A. R. Clarke.** (2007). Enzymatic properties of the lactate dehydrogenase enzyme from *Plasmodium falciparum*. FEBS J **274:**2738-48.

120. **Wickramasinghe, S. R., K. A. Inglis, J. E. Urch, S. Muller, D. M. van Aalten, and A. H. Fairlamb.** (2006). Kinetic, inhibition and structural studies on 3-oxoacyl-ACP reductase from *Plasmodium falciparum*, a key enzyme in fatty acid biosynthesis. Biochem J **393:**447-57.

121. **Ordman, A. B., and S. Kirkwood.** (1977). UDPglucose dehydrogenase. Kinetics and their mechanistic implications. Biochim Biophys Acta **481:**25-32.

122. **Bradbury, S. L., and W. B. Jakoby.** (1971). Ordered binding of substrates to yeast aldehyde dehydrogenase. J Biol Chem **246:**1834-40.

123. **Burger, E., and H. Gorisch.** (1981). Patterns of product inhibition of a bifunctional dehydrogenase; L-histidinol:NAD^+ oxidoreductase. Eur J Biochem **116:**137-42.

124. **Roberts, D. L., F. E. Frerman, and J. J. Kim.** (1996). Three-dimensional structure of human electron transfer flavoprotein to 2.1-A resolution. Proc Natl Acad Sci U S A **93:**14355-60.

125. **Sato, K., Y. Nishina, and K. Shiga.** (2003). Purification of electron-transferring flavoprotein from *Megasphaera elsdenii* and binding of additional FAD with an unusual absorption spectrum. J Biochem **134:**719-29.

126. **Hetzel, M., M. Brock, T. Selmer, A. J. Pierik, B. T. Golding, and W. Buckel.** (2003). Acryloyl-CoA reductase from *Clostridium propionicum*. An enzyme complex of propionyl-CoA dehydrogenase and electron-transferring flavoprotein. Eur J Biochem **270:**902-10.

127. **Herrmann, G.** (2008). Enzymes of two clostridial amino-acid fermentation pathways. Marburg University, Marburg.

128. **Williamson, G., and P. C. Engel.** (1984). Butyryl-CoA dehydrogenase from *Megasphaera elsdenii*. Specificity of the catalytic reaction. Biochem J **218:**521-9.

129. **Diez-Gonzalez, F., J. B. Russell, and J. B. Hunter.** (1997). NAD-independent lactate and butyryl-CoA dehydrogenases of *Clostridium acetobutylicum* P262. Curr Microbiol **34:**162-6.

130. **ROSE, I. A.** (1955). Acetate kinase of bacteria (acetokinase). Methods in Enzymology **1:**591-595.

131. **Stadtman, E. R.** (1952). The net enzymatic synthesis of acetyl coenzyme A. J Biol Chem **196:**535-46.

132. **Stadtman, E. R., and H. A. Barker.** (1950). Fatty acid synthesis by enzyme preparations of *Clostridium kluyveri*. VI. Reactions of acyl phosphates. J Biol Chem **184:**769-93.

133. **Ellman, G. L.** (1959). Tissue sulfhydryl groups. Arch Biochem Biophys **82:**70-7.

134. **Vadali, R. V., G. N. Bennett, and K. Y. San.** (2004). Cofactor engineering of intracellular CoA/acetyl-CoA and its effect on metabolic flux redistribution in *Escherichia coli*. Metab Eng **6:**133-9.

135. **Sanchez, A. M., G. N. Bennett, and K. Y. San.** (2005). Novel pathway engineering design of the anaerobic central metabolic pathway in *Escherichia coli* to increase succinate yield and productivity. Metab Eng **7:**229-39.

136. **Lin, H., G. N. Bennett, and K. Y. San.** (2005). Metabolic engineering of aerobic succinate production systems in *Escherichia coli* to improve process productivity and achieve the maximum theoretical succinate yield. Metab Eng **7:**116-27.

137. **Deng, M. D., D. K. Severson, A. D. Grund, S. L. Wassink, R. P. Burlingame, A. Berry, J. A. Running, C. A. Kunesh, L. Song, T. A. Jerrell, and R. A. Rosson.** (2005). Metabolic engineering of *Escherichia coli* for industrial production of glucosamine and N-acetylglucosamine. Metab Eng **7:**201-14.

138. **Jiang, M., S. W. Liu, J. F. Ma, K. Q. Chen, L. Yu, F. F. Yue, B. Xu, and P. Wei.** (2010). Effect of growth phase feeding strategies on succinate production by metabolically engineered *Escherichia coli*. Appl Environ Microbiol **76:**1298-300.

139. **Veit, A., T. Polen, and V. F. Wendisch.** (2007). Global gene expression analysis of glucose overflow metabolism in *Escherichia coli* and reduction of aerobic acetate formation. Appl Microbiol Biotechnol **74:**406-21.

140. **Vemuri, G. N., E. Altman, D. P. Sangurdekar, A. B. Khodursky, and M. A. Eiteman.** (2006). Overflow metabolism in *Escherichia coli* during steady-state growth: Transcriptional regulation and effect of the redox ratio. Appl Environ Microbiol **72:**3653-61.

141. **Noronha, S. B., H. J. Yeh, T. F. Spande, and J. Shiloach.** (2000). Investigation of the TCA cycle and the glyoxylate shunt in *Escherichia coli* BL21 and JM109 using (13)C-NMR/MS. Biotechnol Bioeng **68:**316-27.

142. **Halpern, Y. S., A. Even-Shoshan, and M. Artman.** (1964). Effect of Glucose on the Utilization of Succinate and the Activity of Tricarboxylic Acid-Cycle Enzymes in *Escherichia coli*. Biochim Biophys Acta **93:**228-36.

143. **van de Walle, M., and J. Shiloach.** (1998). Proposed mechanism of acetate accumulation in two recombinant *Escherichia coli* strains during high density fermentation. Biotechnol Bioeng **57:**71-8.

144. **LaPorte, D. C., K. Walsh, and D. E. Koshland, Jr.** (1984). The branch point effect. Ultrasensitivity and subsensitivity to metabolic control. J Biol Chem **259:**14068-75.

145. **Lin, H., N. M. Castro, G. N. Bennett, and K. Y. San.** (2006). Acetyl-CoA synthetase overexpression in *Escherichia coli* demonstrates more efficient acetate assimilation and lower acetate accumulation: A potential tool in metabolic engineering. Appl Microbiol Biotechnol **71:**870-4.

146. **el-Mansi, E. M., and W. H. Holms.** (1989). Control of carbon flux to acetate excretion during growth of *Escherichia coli* in batch and continuous cultures. J Gen Microbiol **135:**2875-83.

147. **El-Mansi, M., A. J. Cozzone, J. Shiloach, and B. J. Eikmanns.** (2006). Control of carbon flux through enzymes of central and intermediary metabolism during growth of *Escherichia coli* on acetate. Curr Opin Microbiol **9:**173-9.

148. **Eiteman, M. A., and E. Altman.** (2006). Overcoming acetate in *Escherichia coli* recombinant protein fermentations. Trends Biotechnol **24:**530-6.

149. **Lin, H. Y., B. Mathiszik, B. Xu, S. O. Enfors, and P. Neubauer.** (2001). Determination of the maximum specific uptake capacities for glucose and oxygen in glucose-limited fed-batch cultivations of *Escherichia coli*. Biotechnol Bioeng **73:**347-57.

150. **Paalme, T., R. Elken, A. Kahru, K. Vanatalu, and R. Vilu.** (1997). The growth rate control in *Escherichia coli* at near to maximum growth rates: The A-stat approach. Antonie Van Leeuwenhoek **71:**217-30.

151. **Varma, A., and B. O. Palsson.** (1994). Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. Appl Environ Microbiol **60:**3724-31.

152. **Korz, D. J., U. Rinas, K. Hellmuth, E. A. Sanders, and W. D. Deckwer.** (1995). Simple fed-batch technique for high cell density cultivation of *Escherichia coli*. J Biotechnol **39:**59-65.

153. **Paalme, T., R. Elken, A. Kahru, K. Vanatalu & R. Vilu.** (1997). The growth rate control in Escherichia coli at near to maximum growth rates: the A-stat approach. Antonie van Leeuwenhoek **71:**217-230.

154. **Vemuri, G. N., M. A. Eiteman, and E. Altman.** (2006). Increased recombinant protein production in *Escherichia coli* strains with overexpressed water-forming NADH oxidase and a deleted ArcA regulatory protein. Biotechnol Bioeng **94:**538-42.

155. **Majewski, R. A., and M. M. Domach.** (1990). Simple constrained-optimization view of acetate overflow in *E. coli*. Biotechnol Bioeng **35:**732-8.

156. **Koh, B. T., U. Nakashimada, M. P. and, and M. G. S. Yap.** (1992). Comparison of acetate inhibition on growth of host and recombinant *E. coli* K12 strains. Biotechnology letters **14:**1115-1118.

157. **Akesson, M., E. N. Karlsson, P. Hagander, J. P. Axelsson, and A. Tocaj.** (1999). On-line detection of acetate formation in *Escherichia coli* cultures using dissolved oxygen responses to feed transients. Biotechnol Bioeng **64:**590-8.

158. **Crabbendam, P. M., O. M. Neijssel, and D. W. Tempest.** (1985). Metabolic and energetic aspects of the growth of *Clostridium butyricum* on glucose in chemostat culture. Arch Microbiol **142:**375-82.

159. **Johnson, D. E., and R. S. Hanson.** (1974). Bacterial citrate synthases: Purification, molecular weight and kinetic mechanism. Biochim Biophys Acta **350:**336-53.

160. **Huang, L., L. N. Gibbins, and C. W. Forsberg.** (1985). Transmembrane pH gradient and membrane potential in *Clostridium acetobutylicum* during growth under acetogenic and solventogenic conditions. Appl Environ Microbiol **50:**1043-7.

161. **Anderson, A. J., and E. A. Dawes.** (1990). Occurrence, metabolism, metabolic role, and industrial uses of bacterial polyhydroxyalkanoates. Microbiol Rev **54:**450-72.

162. **Senior, P. J., and E. A. Dawes.** (1971). The role and regulation of poly-beta-hydroxybutyrate synthesis in *Azotobacter beijerinckii*. Biochem J **123:**29P.

163. **Bermejo, L. L., N. E. Welker, and E. T. Papoutsakis.** (1998). Expression of *Clostridium acetobutylicum* ATCC 824 genes in *Escherichia coli* for acetone production and acetate detoxification. Appl Environ Microbiol **64:**1079-85.

164. **Jones, D. T., and D. R. Woods.** (1986). Acetone-butanol fermentation revisited. Microbiol Rev **50:**484-524.

165. **Andersch, W., H. Bahl, and G. Gottschalk.** (1983). Level of Enzymes Involved in Acetate, Butyrate, Acetone and Butanol Formation by *Clostridium acetobutylicum*. Eur J Appl Microbiol Biotechnol **18:**327-332.

166. **Janssen, H., C. Doring, A. Ehrenreich, B. Voigt, M. Hecker, H. Bahl, and R. J. Fischer.** (2010). A proteomic and transcriptional view of acidogenic and solventogenic steady-state cells of *Clostridium acetobutylicum* in a chemostat culture. Appl Microbiol Biotechnol **87:**2209-26.

167. **Vanderwinkel, E., P. Furmanski, H. C. Reeves, and S. J. Ajl.** (1968). Growth of *Escherichia coli* on fatty acids: Requirement for coenzyme A transferase activity. Biochem Biophys Res Commun **33:**902-8.

168. **Xu, Y., R. J. Heath, Z. Li, C. O. Rock, and S. W. White.** (2001). The FadR.DNA complex. Transcriptional control of fatty acid metabolism in *Escherichia coli*. J Biol Chem **276:**17373-9.

169. **He, X. Y., S. Y. Yang, and H. Schulz.** (1992). Inhibition of enoyl-CoA hydratase by long-chain L-3-hydroxyacyl-CoA and its possible effect on fatty acid oxidation. Arch Biochem Biophys **298:**527-31.

170. **Pawar, S., and H. Schulz.** (1981). The Structure of the Multienzyme Complex of Fatty Acid Oxidation from *Escherichia coli*. J Biol Chem. **256:**3894- 99.

171. **Leonardo, M. R., Y. Dailly, and D. P. Clark.** (1996). Role of NAD in regulating the adhE gene of *Escherichia coli*. J Bacteriol **178:**6013-8.

172. **Andersen, K. B., and K. von Meyenburg.** (1977). Charges of nicotinamide adenine nucleotides and adenylate energy charge as regulatory parameters of the metabolism in *Escherichia coli*. J Biol Chem **252:**4151-6.

173. **de Graef, M. R., S. Alexeeva, J. L. Snoep, and M. J. Teixeira de Mattos.** (1999). The steady-state internal redox state (NADH/NAD) reflects the external redox state and is correlated with catabolic adaptation in *Escherichia coli*. J Bacteriol **181:**2351-7.

174. **Hedderich, R., and L. Forzi.** (2005). Energy-converting [NiFe] hydrogenases: More than just H_2 activation. J Mol Microbiol Biotechnol **10:**92-104.

175. **Friedrich, T., and D. Scheide.** (2000). The respiratory complex I of bacteria, archaea and eukarya and its module common with membrane-bound multisubunit hydrogenases. FEBS Lett **479:**1-5.

176. **Karen, T., V. Sanchez-Torres, T. K. Wood, A. Trchounian.** (2010). *Escherichia coli* hydrogenase activity and H_2 production under glycerol fermentation at a low pH. International Journal of Hydrogen Energy **36:**10.

177. **Rieu-Lesme, F., C. Dauga, G. Fonty, and J. Dore.** (1998). Isolation from the rumen of a new acetogenic bacterium phylogenetically closely related to *Clostridium difficile*. Anaerobe **4:**89-94.

178. **Ragsdale, S. W., and E. Pierce.** (2008). Acetogenesis and the Wood-Ljungdahl pathway of CO(2) fixation. Biochim Biophys Acta **1784:**1873-98.

179. **Schmidt, S., E. Biegel, and V. Muller.** (2009). The ins and outs of Na($^+$) bioenergetics in *Acetobacterium woodii*. Biochim Biophys Acta **1787:**691-6.

180. **Leclerc, M., A. Bernalier, G. Donadille, and M. Lelait.** (1997). H_2/CO_2 metabolism in acetogenic bacteria isolated from the human colon. Anaerobe **3:**307-15.

181. **Karasawa, T., S. Ikoma, K. Yamakawa, and S. Nakamura.** (1995). A defined growth medium for Clostridium difficile. Microbiology **141 (Pt 2):**371-5.

182. **Furdui, C., and S. W. Ragsdale.** (2000). The role of pyruvate ferredoxin oxidoreductase in pyruvate synthesis during autotrophic growth by the Wood-Ljungdahl pathway. J Biol Chem **275:**28494-9.

183. **Furdui, C., and S. W. Ragsdale.** (2002). The roles of coenzyme A in the pyruvate:ferredoxin oxidoreductase reaction mechanism: rate enhancement of electron transfer from a radical intermediate to an iron-sulfur cluster. Biochemistry **41:**9921-37.

184. **Siriwongrungson, V., R. J. Zeng, and I. Angelidaki.** (2007). Homoacetogenesis as the alternative pathway for H_2 sink during thermophilic anaerobic degradation of butyrate under suppressed methanogenesis. Water Res **41:**4204-10.

App.1: List of primers used in the current work

thia1-CD1059-s	AAGCTCTTCAATGAGAGAAGTAGTAATTGCCAGTG
thia1-CD1059-as	TTGCTCTTCTCCCTCTCTTAACTATTAAAGTAGTTCCCATTC
hbd-CD1058-s	AAGCTCTTCAATGAAATTAGCTGTAATAGGTAGTGGAAC
hbd-CD1058-as	TTGCTCTTCTCCCTTTATTATAATCATAGAATCCTATCTTAGT
crt2-CD1057-s	AAGCTCTTCAATGAGTACAAGTGATGTTAAAGTTTATGAG
crt2-CD1057-as	TTGCTCTTCTCCCCCCTTTTATAAAGTTAGCTTCTCTCTTTTC
bcd2-CD1054-s	AAGCTCTTCAATGGATTTAAATTCTAAAAAATATCAGATGCT
bcd2-CD1054-as	TTGCTCTTCTCCCTTTTAATAGTTTTCCTGAAATAACCATTCT
etfB2-CD1055-s	AAGCTCTTCAATGGAGGAAGGATTTATGAATATAGTCG
etfB2-CD1055-as	TTGCTCTTCTCCCTATGATATACTTCTCTTTTAATTTATCTATG
etfA2-CD1056-s	AAGCTCTTCAATGGGTAACGTTTTAGTAGTAATAGAACAAAG
etfA2-CD1056-as	TTGCTCTTCTCCCGTTAGCTAAAACTTCACCTTTTTCTTTTGC
buk-CD0113-s	AAGCTCTTCAATGAGCAAAATATTTAAAATCTTAACAATAAATCCTGG
bukCD0113-as	AAGCTCTTCTCCCGTTATCATAAACTTGAGCCTCTTCTTCA
pbt-CD0112-s	AAGCTCTTCAATGAGAAGTTTTGAAGAAGTAATTAAGTTTGC
pbt-CD0112-mut1	AAGCTCTTCTCGTCCTGCTACAGGATGA
pbt-CD0112-mut2	AAGCTCTTCAACGAGCTGATATATTATTAGCCCC
pbt-CD0112-as	AAGCTCTTCTCCCTGCCTTTGCTGCCATTAAAACA
bcd_CA_C2711_s	AAGCTCTTCAATGGATTTTAATTTAACAAGAGAACAAGAATTAGTAAG
bcd_CA_C2711_as	AAGCTCTTCTCCCTCTAAAAATTTTTCCTGAAATAACTAATTTCTGAAC
etfB_CA_C2710_s	AAGCTCTTCAATGAATATAGTTGTTTGTTTAAAACAAGTTCC
etfB_CA_C2710_as	AAGCTCTTCTCCCAATATAGTGTTCTTCTTTTAATTTTGAGACAAC
etfA_CA_C2709_s	AAGCTCTTCAAGATGGAGAATTACAAAAGGTATCATTGG
etfA_CA_C2709_as	AAGCTCTTCTCCCATTATTAGCAGCTTTAACTTGAGC
etfA_ CA_C2709_SO1	ATGAATAAAGCAGATTACAAGGGCGTATGGGTGTTTGCTGAACAAAG
etfA_ CA_C2709_SO2	TCTCTTTGTTCAGCAAACACCCATACGCCCTTGTAATCTGCTTTATT

S, refer to sense primer (forward); as, is antisense primer (reverse); mut, primer to insert a silent mutation.

App.2: List of cloning, expression and fusion plasmids used in the current study

Plasmid	Special characteristics	Source
Cloning plasmids		
pE_thiA$_1$	Kan^r: with PCR fragment for *C. difficile thiA$_1$*(CD1059)	R. Fischer
pE_hbd	Kan^r: with PCR fragment for *C. difficile hbd* (CD1058)	R. Fischer
pE_crt$_2$	Kan^r: with PCR fragment for *C. difficile crt$_2$* (CD1057)	R. Fischer
pE_bcdABC$_2$	Kan^r: with fused fragments for *C. difficile bcd$_2$* (CD1054) and the two electron transfer subunits *etfB$_2$* (CD1055), and *etfA$_2$* (CD1056)	R. Fischer
pE-bcd	Kan^r: with PCR fragment for *bcd* (CA_C2711) from *C. acetobutylicum*	This study
pE-etfB	Kan^r: with PCR fragment for *etfB* (CA_C2710) from *C. acetobutylicum*	This study
pE-etfA	Kan^r: with PCR fragment for *etfA* (CA_C2709) from *C. acetobutylicum*	This study
pE_bcdABC	Kan^r: with fused fragments *bcd* (CA_C2711) and the tow electron transfer subunits *etfB* (CA_C2710), and *etfA* (CA_C2709) from *C. acetobutylicum*	This study
pE_pbt	Kan^r: with PCR fragment for *C. difficile* pbt (CD0112)	This study
pE_buk	Kan^r: with PCR fragment for *C. difficile* buk (CD0113)	This study
Expression plasmids		
pASG.wt_thiA$_1$	Amp^r: Expression vector for the wild type *thiA$_1$*	This study
pASG.3_thiA$_1$	Amp^r: Expression vector for *thiA$_1$* with *Strep-Tag II* attached with *C*-terminal	R. Fischer
pASG.5_thiA$_1$	Amp^r: Expression vector for *thiA$_1$* with *Strep-Tag II* attached with *N*-terminal	R. Fischer
pASG.wt_hbd	Amp^r: Expression vector for the wild type *hbd*	This study
pASG.3_hbd	Amp^r: Expression vector for *hbd* with *Strep-Tag* attached with *C*-terminal	R. Fischer
pASG.5_hbd	Amp^r: Expression vector for *hbd* with *Strep-Tag II* attached with *N*-terminal	R. Fischer
pASG.wt_crt$_2$	Amp^r: Expression vector for the wild type *crt$_2$*	This study
pASG.3_crt$_2$	Amp^r: Expression vector for *crt$_2$* with *Strep-Tag* attached with *C*-terminal	R. Fischer
pASG.5_crt$_2$	Amp^r: Expression vector for *crt$_2$* with *Strep-Tag II* attached with *N*-terminal	R. Fischer

pASG.wt_pbt	Amp^r: Expression vector for the wild type *pbt*	This study
pASG.3_pbt	Amp^r: Expression vector for *pbt* with *Strep-Tag II* attached with *C*-terminal	This study
pASG.5_pbt	Amp^r: Expression vector for *pbt* with *Strep-Tag II* attached with *N*-terminal	This study
pASG.wt_buk	Amp^r: Expression vector for the wild type buk	This study
pASG.3_buk	Amp^r: Expression vector for *buk* with *Strep-Tag II* attached with *C*-terminal	This study
pASG.5_buk	Amp^r: Expression vector for *buk* with *Strep-Tag II* attached with *N*-terminal	This study
pASG.3_bcdABC$_2$	Amp^r: Expression vector with fused *bcd*$_2$, *etfB*$_2$, and *etfA*$_2$ from *C. difficile* with *Strep-Tag II* attached with *C*-terminal	R. Fischer
pASG.5_bcdABC$_2$	Amp^r: Expression vector with fused *bcd*$_2$, *etfB*$_2$, and *etfA*$_2$ from *C. difficile* with *Strep-Tag II* attached with *N*-terminal	R. Fischer
pASG.3_bcdABC	Amp^r: Expression vector with fused *bcd*, *etfB*, and *etfA* from *C. acetobutylicum* with *Strep-Tag II* attached with *C*-terminal	This study
pASG.5_bcdABC	Amp^r: Expression vector with fused *bcd*, *etfB*, and *etfA* from *C. acetobutylicum* with *Strep-Tag II* attached with *N*-terminal	This study
pASG.3_bcdBCA	Amp^r: Expression vector with fused *bcd*, *etfB*, and *etfA* from *C. acetobutylicum* with *Strep-Tag II* attached with *C*-terminal	This study
pASG.5_bcdBCA	Amp^r: Expression vector with fused *etfB*, *etfA*, and *bcd* from *C. acetobutylicum* with *Strep-Tag II* attached with *N*-terminal	This study
pASG.3_bcdCAB	Amp^r: Expression vector with fused *etfB*, *etfA*, and *bcd* from *C. acetobutylicum* with *Strep-Tag II* attached with *C*-terminal	This study
pASG.5_bcdCAB	Amp^r: Expression vector with fused *etfA*, *bcd*, and *etfB* from *C. acetobutylicum* with *Strep-Tag II* attached with *N*-terminal	This study
pASG.wt_pbt-buk	Amp^r: Expression vector with fused *pbt* and *buk*.	This study
pASG.wt_thiA$_1$*-hbd - crt*$_2$	Amp^r: Expression vector with fused *thiA*$_1$, *hbd*, and *crt*$_2$.	This study
pASG.wt_6 genes$_D$	Amp^r: Expression vector with fused *thiA*$_1$, *hbd*, *crt*$_2$, and *bcdABC*$_2$	This study
pASG.wt_6 genes$_{AC}$	Amp^r: Expression vector with fused *thiA*$_1$, *hbd*, *crt*$_2$, and *bcdABC*	This study
pASG.wt_BUT$_D$	Amp^r: Expression vector with fused butyrate biosynthesis	This study

	pathway genes (*thiA$_1$*, *hbd*, *crt$_2$*, *bcdABC$_2$*, *pbt*, and *buk*)	
pASG.wt_BUT$_{AC}$	Amp^r: Expression vector with fused butyrate biosynthesis pathway genes (*thiA$_1$*, *hbd*, *crt$_2$*, *bcdABC*, *pbt*, and *buk*)	This study
Fusion plasmids		
pFF.rbs_bcd$_2$.rbs_etfB$_2$.rbs_etfA$_2$	Cluster I *bcdABC$_2$*	This study
pFF.rbs_bcd	Cluster I´ *bcdABC*	This study
pFFc_etfB		This study
pFFc_etfA		This study
pE_bcd.rbs_etfB		This study
pFF.rbs_bcd.rbs_etfB		This study
pE_bcd.rbs_etfB.rbs_etfA		This study
pFF.rbs_bcd.rbs_etfB.rbs_etfA		This study
pFF.rbs_thiA$_1$	Cluster II *thiA$_1$-hbd-crt$_2$*	This study
pFFc_hbd		This study
pFFc_crt$_2$		This study
pE_thiA$_1$.rbs_hbd		This study
pFF.rbs_thiA$_1$.rbs_hbd		This study
pE_thiA$_1$.rbs_hbd.rbs_crt$_2$		This study
pFFc_thiA$_1$.rbs_hbd.rbs_crt$_2$		This study
pFF.rbs_pbt	Cluster III *Pbt-buk*	This study
pFFc_buk		This study
pE_pbt.rbs_buk		This study
pFFc_pbt.rbs_buk		This study
pE_bcd$_2$.rbs_etfB$_2$.rbs_etfA$_2$.rbs_thiA$_1$.rbs_hbd.rbs_crt$_2$	Fused cluster I & II together (*6 genes$_D$*)	This study
pFF.rbs_bcd$_2$.rbs_etfB$_2$.rbs_etfA$_2$.rbs_thiA$_1$.rbs_hbd.rbs_crt$_2$		This study
pE_bcd.rbs_etfB.rbs_etfA.rbs_thiA$_1$.rbs_hbd.rbs_crt$_2$	Fused cluster I´ & II together (*6 genes$_{AC}$*)	This study
pFF.rbs_bcd.rbs_etfB.rbs_etfA.rbs_thiA$_1$.rbs_hbd.rbs_crt$_2$		This study
pE_bcd$_2$.rbs_etfB$_2$.rbs_etfA$_2$.rbs_thiA$_1$.rbs_hbd.rbs_crt$_2$.rbs_pbt.rbs_buk	Fused cluster I, II & III together (*pE-BUT$_D$*)	This study
pE_bcd.rbs_etfB.rbs_etfA.rbs_thiA$_1$.rbs_hbd.rbs_crt$_2$.rbs_pbt.rbs_buk	Fused cluster I´, II & III together (*pE-BUT$_{AC}$*)	This study

```
>pE-CD1054_insert
      1 CGTCTCCAAT GGATTTAAAT TCTAAAAAAT ATCAGATGCT TAAAGAGCTA TATGTAAGCT
     61 TCGCTGAAAA TGAAGTTAAA CCTTTAGCAA CAGAACTTGA TGAAGAAGAA AGATTTCCTT
    121 ATGAAACAGT GGAAAAAATG GCAAAAGCAG GAATGATGGG TATACCATAT CCAAAAGAAT
    181 ATGGTGGAGA AGGTGGAGAC ACTGTAGGAT ATATAATGGC AGTTGAAGAA TTGTCTAGAG
    241 TTTGTGGTAC TACAGGAGTT ATATTATCAG CTCATACATC TCTTGGCTCA TGGCCTATAT
    301 ATCAATATGG TAATGAAGAA CAAAAACAAA AATTCTTAAG ACCACTAGCA AGTGGAGAAA
    361 AATTAGGAGC ATTTGGTCTT ACTGAGCCTA ATGCTGGTAC AGATGCGTCT GGCCAACAAA
    421 CAACTGCTGT TTTAGACGGG GATGAATACA TACTTAATGG CTCAAAAATA TTTATAACAA
    481 ACGCAATAGC TGGTGACATA TATGTAGTAA TGGCAATGAC TGATAAATCT AAGGGGAACA
    541 AAGGAATATC AGCATTTATA GTTGAAAAAG GAACTCCTGG GTTTAGCTTT GGAGTTAAAG
    601 AAAAGAAAAT GGGTATAAGA GGTTCAGCTA CGAGTGAATT AATATTTGAG GATTGCAGAA
    661 TACCTAAAGA AAATTTACTT GGAAAAGAAG GTCAAGGATT TAAGATAGCA ATGTCTACTC
    721 TTGATGGTGG TAGAATTGGT ATAGCTGCAC AAGCTTTAGG TTTAGCACAA GGTGCTCTTG
    781 ATGAAACTGT TAAATATGTA AAAGAAAGAG TACAATTTGG TAGACCATTA TCAAAATTCC
    841 AAAATACACA ATTCCAATTA GCTGATATGG AAGTTAAGGT ACAAGCGGCT AGACACCTTG
    901 TATATCAAGC AGCTATAAAT AAAGACTTAG GAAAACCTTA TGGAGTAGAA GCAGCAATGG
    961 CAAAATTATT TGCAGCTGAA ACAGCTATGG AAGTTACTAC AAAAGCTGTA CAACTTCATG
   1021 GAGGATATGG ATACACTCGT GACTATCCAG TAGAAAGAAT GATGAGAGAT GCTAAGATAA
   1081 CTGAAATATA TGAAGGAACT AGTGAAGTTC AAAGAATGGT TATTTCAGGA AAACTATTAA
   1141 AAGGGAGGAG ACG
```

App.3: The nucleotide sequence of *bcd*$_2$ cloned from *C. difficile* DSMZ 1296^T. The gray highlighted sequences refer to the *Esp3I* restriction site. The stop codon was replaced with *GGG*.

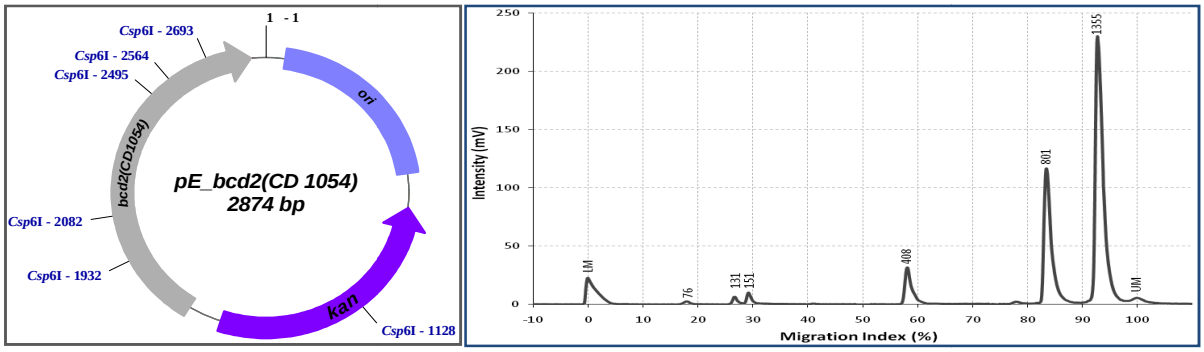
```
pE_CD1054-insert  CGTCTCCAAT GGATTTAAAT TCTAAAAAAT ATCAGATGCT TAAAGAGCTA
     CD1054-bcd2  --------AT GGATTTAAAT TCTAAAAAAT ATCAGATGCT TAAAGAGCTA

pE_CD1054-insert  TATGTAAGCT TCGCTGAAAA TGAAGTTAAA CCTTTAGCAA CAGAACTTGA
     CD1054-bcd2  TATGTAAGCT TCGCTGAAAA TGAAGTTAAA CCTTTAGCAA CAGAACTTGA

pE_CD1054-insert  TGAAGAAGAA AGATTTCCTT ATGAAACAGT GGAAAAAATG GCAAAAGCAG
     CD1054-bcd2  TGAAGAAGAA AGATTTCCTT ATGAAACAGT GGAAAAAATG GCAAAAGCAG

pE_CD1054-insert  GAATGATGGG TATACCATAT CCAAAAGAAT ATGGTGGAGA AGGTGGAGAC
     CD1054-bcd2  GAATGATGGG TATACCATAT CCAAAAGAAT ATGGTGGAGA AGGTGGAGAC

pE_CD1054-insert  ACTGTAGGAT ATATAATGGC AGTTGAAGAA TTGTCTAGAG TTTGTGGTAC
     CD1054-bcd2  ACTGTAGGAT ATATAATGGC AGTTGAAGAA TTGTCTAGAG TTTGTGGTAC

pE_CD1054-insert  TACAGGAGTT ATATTATCAG CTCATACATC TCTTGGCTCA TGGCCTATAT
     CD1054-bcd2  TACAGGAGTT ATATTATCAG CTCATACATC TCTTGGCTCA TGGCCTATAT

pE_CD1054-insert  ATCAATATGG TAATGAAGAA CAAAAACAAA AATTCTTAAG ACCACTAGCA
     CD1054-bcd2  ATCAATATGG TAATGAAGAA CAAAAACAAA AATTCTTAAG ACCACTAGCA

pE_CD1054-insert  AGTGGAGAAA AATTAGGAGC ATTTGGTCTT ACTGAGCCTA ATGCTGGTAC
     CD1054-bcd2  AGTGGAGAAA AATTAGGAGC ATTTGGTCTT ACTGAGCCTA ATGCTGGTAC

pE_CD1054-insert  AGATGCGTCT GGCCAACAAA CAACTGCTGT TTTAGACGGG GATGAATACA
     CD1054-bcd2  AGATGCGTCT GGCCAACAAA CAACTGCTGT TTTAGACGGG GATGAATACA

pE_CD1054-insert  TACTTAATGG CTCAAAAATA TTTATAACAA ACGCAATAGC TGGTGACATA
     CD1054-bcd2  TACTTAATGG CTCAAAAATA TTTATAACAA ACGCAATAGC TGGTGACATA

pE_CD1054-insert  TATGTAGTAA TGGCAATGAC TGATAAATCT AAGGGGAACA AAGGAATATC
     CD1054-bcd2  TATGTAGTAA TGGCAATGAC TGATAAATCT AAGGGGAACA AAGGAATATC

pE_CD1054-insert  AGCATTTATA GTTGAAAAAG GAACTCCTGG GTTTAGCTTT GGAGTTAAAG
     CD1054-bcd2  AGCATTTATA GTTGAAAAAG GAACTCCTGG GTTTAGCTTT GGAGTTAAAG

pE_CD1054-insert  AAAAGAAAAT GGGTATAAGA GGTTCAGCTA CGAGTGAATT AATATTTGAG
     CD1054-bcd2  AAAAGAAAAT GGGTATAAGA GGTTCAGCTA CGAGTGAATT AATATTTGAG
```

```
pE_CD1054-insert   GATTGCAGAA TACCTAAAGA AAATTTACTT GGAAAAGAAG GTCAAGGATT
     CD1054-bcd2   GATTGCAGAA TACCTAAAGA AAATTTACTT GGAAAAGAAG GTCAAGGATT

pE_CD1054-insert   TAAGATAGCA ATGTCTACTC TTGATGGTGG TAGAATTGGT ATAGCTGCAC
     CD1054-bcd2   TAAGATAGCA ATGTCTACTC TTGATGGTGG TAGAATTGGT ATAGCTGCAC

pE_CD1054-insert   AAGCTTTAGG TTTAGCACAA GGTGCTCTTG ATGAAACTGT TAAATATGTA
     CD1054-bcd2   AAGCTTTAGG TTTAGCACAA GGTGCTCTTG ATGAAACTGT TAAATATGTA

pE_CD1054-insert   AAAGAAAGAG TACAATTTGG TAGACCATTA TCAAAATTCC AAAATACACA
     CD1054-bcd2   AAAGAAAGAG TACAATTTGG TAGACCATTA TCAAAATTCC AAAATACACA

pE_CD1054-insert   ATTCCAATTA GCTGATATGG AAGTTAAGGT ACAAGCGGCT AGACACCTTG
     CD1054-bcd2   ATTCCAATTA GCTGATATGG AAGTTAAGGT ACAAGCGGCT AGACACCTTG

pE_CD1054-insert   TATATCAAGC AGCTATAAAT AAAGACTTAG GAAAACCTTA TGGAGTAGAA
     CD1054-bcd2   TATATCAAGC AGCTATAAAT AAAGACTTAG GAAAACCTTA TGGAGTAGAA

pE_CD1054-insert   GCAGCAATGG CAAAATTATT TGCAGCTGAA ACAGCTATGG AAGTTACTAC
     CD1054-bcd2   GCAGCAATGG CAAAATTATT TGCAGCTGAA ACAGCTATGG AAGTTACTAC

pE_CD1054-insert   AAAAGCTGTA CAACTTCATG GAGGATATGG ATACACTCGT GACTATCCAG
     CD1054-bcd2   AAAAGCTGTA CAACTTCATG GAGGATATGG ATACACTCGT GACTATCCAG

pE_CD1054-insert   TAGAAAGAAT GATGAGAGAT GCTAAGATAA CTGAAATATA TGAAGGAACT
     CD1054-bcd2   TAGAAAGAAT GATGAGAGAT GCTAAGATAA CTGAAATATA TGAAGGAACT

pE_CD1054-insert   AGTGAAGTTC AAAGAATGGT TATTTCAGGA AAACTATTAA AAGGGAGGAG
     CD1054-bcd2   AGTGAAGTTC AAAGAATGGT TATTTCAGGA AAACTATTAA AATAG-----

pE_CD1054-insert   ACG
     CD1054-bcd2   ---
```

App.4: Sequence alignment for the cloned *bcd*$_2$ from *C. difficile* DSMZ 1296^T and the genome derived genes from strain 630 [131] (GeneBank accession number CD1054). No mutation was observed.

```
      pE_CD1054   MDLNSKKYQM LKELYVSFAE NEVKPLATEL DEEERFPYET VEKMAKAGMM
           Bcd2   MDLNSKKYQM LKELYVSFAE NEVKPLATEL DEEERFPYET VEKMAKAGMM

      pE_CD1054   GIPYPKEYGG EGGDTVGYIM AVEELSRVCG TTGVILSAHT SLGSWPIYQY
           Bcd2   GIPYPKEYGG EGGDTVGYIM AVEELSRVCG TTGVILSAHT SLGSWPIYQY

      pE_CD1054   GNEEQKQKFL RPLASGEKLG AFGLTEPNAG TDASGQQTTA VLDGDEYILN
           Bcd2   GNEEQKQKFL RPLASGEKLG AFGLTEPNAG TDASGQQTTA VLDGDEYILN

      pE_CD1054   GSKIFITNAI AGDIYVVMAM TDKSKGNKGI SAFIVEKGTP GFSFGVKEKK
           Bcd2   GSKIFITNAI AGDIYVVMAM TDKSKGNKGI SAFIVEKGTP GFSFGVKEKK

      pE_CD1054   MGIRGSATSE LIFEDCRIPK ENLLGKEGQG FKIAMSTLDG GRIGIAAQAL
           Bcd2   MGIRGSATSE LIFEDCRIPK ENLLGKEGQG FKIAMSTLDG GRIGIAAQAL

      pE_CD1054   GLAQGALDET VKYVKERVQF GRPLSKFQNT QFQLADMEVK VQAARHLVYQ
           Bcd2   GLAQGALDET VKYVKERVQF GRPLSKFQNT QFQLADMEVK VQAARHLVYQ

      pE_CD1054   AAINKDLGKP YGVEAAMAKL FAAETAMEVT TKAVQLHGGY GYTRDYPVER
           Bcd2   AAINKDLGKP YGVEAAMAKL FAAETAMEVT TKAVQLHGGY GYTRDYPVER

      pE_CD1054   MMRDAKITEI YEGTSEVQRM VISGKLLKG
           Bcd2   MMRDAKITEI YEGTSEVQRM VISGKLLK.
```

App.5: Amino acid alignment for the cloned *bcd*$_2$ from *C. difficile* DSMZ 1296^T and the genome derived genes from strain 630 [131]. No mutation was observed.



Expected Fragments size (bp)	69	129	150	413	804	1309

App.6: Analysis of *pE_bcd$_2$ (CD1054)* with MultiNA after digestion with *Csp61* restriction enzyme. On the right hand, there is electropherogram for the plasmid. On the left hand, there is pDRAW file shows the recognition sites of *Csp61* as well as the position of the sequencing primers.

```
>pE_CD1055-insert
      1 CGTCTCCAAT GGAGGAAGGA TTTATGAATA TAGTCGTTTG TATAAAACAA GTTCCAGATA
     61 CAACAGAAGT TAAACTAGAT CCTAATACAG GTACTTTAAT TAGAGATGGA GTACCAAGTA
    121 TAATAAACCC TGATGATAAA GCAGGTTTAG AAGAAGCTAT AAAATTAAAA GAAGAAATGG
    181 GTGCTCATGT AACTGTTATA ACAATGGGAC CTCCTCAAGC AGATATGGCT TTAAAAGAAG
    241 CTTTAGCAAT GGGTGCAGAT AGAGGTATAT TATTAACAGA TAGAGCATTT GCAGGTGCTG
    301 ATACTTGGGC AACTTCATCA GCATTAGCAG GAGCATTAAA AAATATAGAT TTTGATATTA
    361 TAATAGCTGG AAGACAGGCG ATAGATGGAG ATACTGCACA AGTTGGACCT CAAATAGCTG
    421 AACATTTAAA TCTTCCATCA ATAACATATG CTGAAGAAAT AAAAACTGAA GGTGAATATG
    481 TATTAGTAAA AAGACAATTT GAAGATTGTT GCCATGACTT AAAAGTTAAA ATGCCATGCC
    541 TTATAACAAC TCTTAAAGAT ATGAACACAC CAAGATACAT GAAAGTTGGA AGAATATATG
    601 ATGCTTTCGA AAATGATGTA GTAGAAACAT GGACTGTAAA AGATATAGAA GTTGACCCTT
    661 CTAATTTAGG TCTTAAAGGT TCTCCAACTA GTGTATTTAA ATCATTTACA AAATCAGTTA
    721 AACCAGCTGG TACAATATAC AATGAAGATG CGAAAACATC AGCTGGAATT ATCATAGATA
    781 AATTAAAAGA GAAGTATATC ATAGGGAGGA GACG
```

App.7: The nucleotide sequence of *etfB$_2$* cloned from *C. difficile* DSMZ 1296^T. The gray highlighted sequences refer to the *Esp3I* restriction site. The stop codon was replaced with *GGG*.

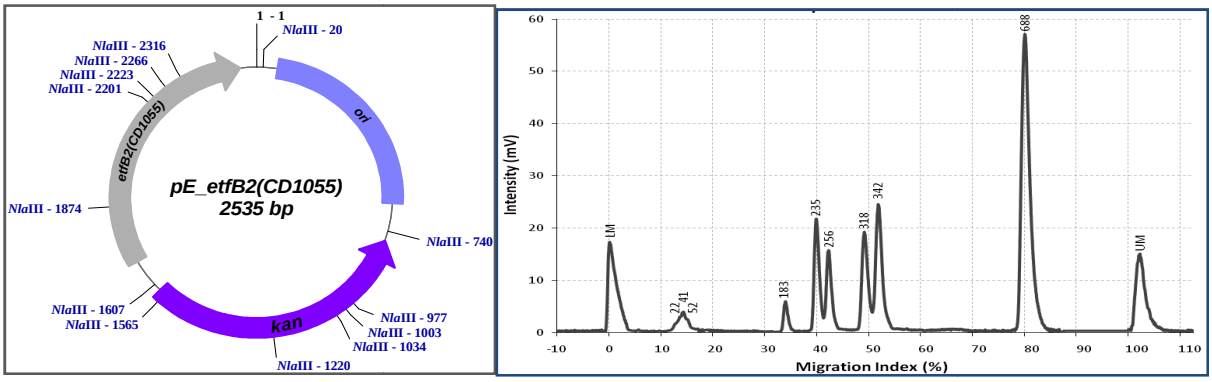
```
pE_CD1055-insert  CGTCTCCAAT GGAGGAAGGA TTTATGAATA TAGTCGTTTG TATAAAACAA
    CD1055-etfB2  --------AT GGAGGAAGGA TTTATGAATA TAGTCGTTTG TATAAAACAA

pE_CD1055-insert  GTTCCAGATA CAACAGAAGT TAAACTAGAT CCTAATACAG GTACTTTAAT
    CD1055-etfB2  GTTCCAGATA CAACAGAAGT TAAACTAGAT CCTAATACAG GTACTTTAAT

pE_CD1055-insert  TAGAGATGGA GTACCAAGTA TAATAAACCC TGATGATAAA GCAGGTTTAG
    CD1055-etfB2  TAGAGATGGA GTACCAAGTA TAATAAACCC TGATGATAAA GCAGGTTTAG

pE_CD1055-insert  AAGAAGCTAT AAAATTAAAA GAAGAAATGG GTGCTCATGT AACTGTTATA
    CD1055-etfB2  AAGAAGCTAT AAAATTAAAA GAAGAAATGG GTGCTCATGT AACTGTTATA

pE_CD1055-insert  ACAATGGGAC CTCCTCAAGC AGATATGGCT TTAAAAGAAG CTTTAGCAAT
    CD1055-etfB2  ACAATGGGAC CTCCTCAAGC AGATATGGCT TTAAAAGAAG CTTTAGCAAT

pE_CD1055-insert  GGGTGCAGAT AGAGGTATAT TATTAACAGA TAGAGCATTT GCAGGTGCTG
    CD1055-etfB2  GGGTGCAGAT AGAGGTATAT TATTAACAGA TAGAGCATTT GCAGGTGCTG

pE_CD1055-insert  ATACTTGGGC AACTTCATCA GCATTAGCAG GAGCATTAAA AAATATAGAT
    CD1055-etfB2  ATACTTGGGC AACTTCATCA GCATTAGCAG GAGCATTAAA AAATATAGAT

pE_CD1055-insert  TTTGATATTA TAATAGCTGG AAGACAGGCG ATAGATGGAG ATACTGCACA
    CD1055-etfB2  TTTGATATTA TAATAGCTGG AAGACAGGCG ATAGATGGAG ATACTGCACA

pE_CD1055-insert  AGTTGGACCT CAAATAGCTG AACATTTAAA TCTTCCATCA ATAACATATG
    CD1055-etfB2  AGTTGGACCT CAAATAGCTG AACATTTAAA TCTTCCATCA ATAACATATG
```

```
pE_CD1055-insert   CTGAAGAAAT AAAAACTGAA GGTGAATATG TATTAGTAAA AAGACAATTT
    CD1055-etfB2   CTGAAGAAAT AAAAACTGAA GGTGAATATG TATTAGTAAA AAGACAATTT

pE_CD1055-insert   GAAGATTGTT GCCATGACTT AAAAGTTAAA ATGCCATGCC TTATAACAAC
    CD1055-etfB2   GAAGATTGTT GCCATGACTT AAAAGTTAAA ATGCCATGCC TTATAACAAC

pE_CD1055-insert   TCTTAAAGAT ATGAACACAC CAAGATACAT GAAAGTTGGA AGAATATATG
    CD1055-etfB2   TCTTAAAGAT ATGAACACAC CAAGATACAT GAAAGTTGGA AGAATATATG

pE_CD1055-insert   ATGCTTTCGA AAATGATGTA GTAGAAACAT GGACTGTAAA AGATATAGAA
    CD1055-etfB2   ATGCTTTCGA AAATGATGTA GTAGAAACAT GGACTGTAAA AGATATAGAA

pE_CD1055-insert   GTTGACCCTT CTAATTTAGG TCTTAAAGGT TCTCCAACTA GTGTATTTAA
    CD1055-etfB2   GTTGACCCTT CTAATTTAGG TCTTAAAGGT TCTCCAACTA GTGTATTTAA

pE_CD1055-insert   ATCATTTACA AAATCAGTTA AACCAGCTGG TACAATATAC AATGAAGATG
    CD1055-etfB2   ATCATTTACA AAATCAGTTA AACCAGCTGG TACAATATAC AATGAAGATG

pE_CD1055-insert   CGAAAACATC AGCTGGAATT ATCATAGATA AATTAAAAGA GAAGTATATC
    CD1055-etfB2   CGAAAACATC AGCTGGAATT ATCATAGATA AATTAAAAGA GAAGTATATC

pE_CD1055-insert   ATAGGGAGGA GACG
    CD1055-etfB2   ATATAA---- ----
```

App.8: Sequence alignment for the cloned *etfB*$_2$ from *C. difficile* DSMZ 1296^T and the genome derived genes from strain 630 [131] (GeneBank accession number CD1055). No mutation was observed.

```
      pE_CD1055   MEEGFMNIVV CIKQVPDTTE VKLDPNTGTL IRDGVPSIIN PDDKAGLEEA
          etfB2   MEEGFMNIVV CIKQVPDTTE VKLDPNTGTL IRDGVPSIIN PDDKAGLEEA

      pE_CD1055   IKLKEEMGAH VTVITMGPPQ ADMALKEALA MGADRGILLT DRAFAGADTW
          etfB2   IKLKEEMGAH VTVITMGPPQ ADMALKEALA MGADRGILLT DRAFAGADTW

      pE_CD1055   ATSSALAGAL KNIDFDIIIA GRQAIDGDTA QVGPQIAEHL NLPSITYAEE
          etfB2   ATSSALAGAL KNIDFDIIIA GRQAIDGDTA QVGPQIAEHL NLPSITYAEE

      pE_CD1055   IKTEGEYVLV KRQFEDCCHD LKVKMPCLIT TLKDMNTPRY MKVGRIYDAF
          etfB2   IKTEGEYVLV KRQFEDCCHD LKVKMPCLIT TLKDMNTPRY MKVGRIYDAF

      pE_CD1055   ENDVVETWTV KDIEVDPSNL GLKGSPTSVF KSFTKSVKPA GTIYNEDAKT
          etfB2   ENDVVETWTV KDIEVDPSNL GLKGSPTSVF KSFTKSVKPA GTIYNEDAKT

      pE_CD1055   SAGIIIDKLK EKYIIG
          etfB2   SAGIIIDKLK EKYII.
```

App.9: Amino acid alignment for the cloned *etfB*$_2$ from *C. difficile* DSMZ 1296^T and the genome derived genes from strain 630 [131]. No mutation was observed.



Expected Fragments (bp)	22+26+31	42+43	50	186	237+239	267	327	345	720

App.10: Analysis of *pE_etfB*$_2$ *(CD1055)* with MultiNA after digestion with *NlaIII* restriction enzyme. On the right hand, there is electropherogram for the plasmid. On the left hand, there is pDRAW file shows the recognition sites of *NlaIII* as well as the position of the sequencing primers.

```
>pE_CD1056- insert
      1 CGTCTCCAAT GGGTAACGTT TTAGTAGTAA TAGAACAAAG AGAAAATGTA ATTCAAACTG
     61 TTTCTTTAGA ATTACTAGGA AAGGCTACAG AAATAGCAAA AGATTATGAT ACAAAAGTTT
    121 CTGCATTACT TTTAGGTAGT AAGGTAGAAG GTTTAATAGA TACATTAGCA CACTATGGTG
    181 CAGATGAGGT AATAGTAGTA GATGATGAAG CTTTAGCAGT GTATACAACT GAACCATATA
    241 CAAAAGCAGC TTATGAAGCA ATAAAAGCAG CTGACCCTAT AGTTGTATTA TTTGGTGCAA
    301 CTTCAATAGG TAGAGATTTA GCGCCTAGAG TTTCTGCTAG AATACATACA GGTCTTACTG
    361 CTGACTGTAC AGGTCTTGCA GTAGCTGAAG ATACAAAATT ATTATTAATG ACAAGACCTG
    421 CCTTTGGTGG AAATATAATG GCAACAATAG TTTGTAAAGA TTTCAGACCT CAAATGTCTA
    481 CAGTTAGACC AGGGGTTATG AAGAAAAATG AACCTGATGA AACTAAAGAA GCTGTAATTA
    541 ACCGTTTCAA GGTAGAATTT AATGATGCTG ATAAATTAGT TCAAGTTGTA CAAGTAATAA
    601 AAGAAGCTAA AAAACAAGTT AAAATAGAAG ATGCTAAGAT ATTAGTTTCT GCTGGACGTG
    661 GAATGGGTGG AAAAGAAAAC TTAGACATAC TTTATGAATT AGCTGAAATT ATAGGTGGAG
    721 AAGTTTCTGG TTCTCGTGCC ACTATAGATG CAGGTTGGTT AGATAAAGCA AGACAAGTTG
    781 GTCAAACTGG TAAAACTGTA AGACCAGACC TTTATATAGC ATGTGGTATA TCTGGAGCAA
    841 TACAACATAT AGCTGGTATG GAAGATGCTG AGTTTATAGT TGCTATAAAT AAAAATCCAG
    901 AAGCTCCAAT ATTTAAATAT GCTGATGTTG GTATAGTTGG AGATGTTCAT AAAGTGCTTC
    961 CAGAACTTAT CAGTCAGTTA AGTGTTGCAA AAGAAAAAGG TGAAGTTTTA GCTAACGGGA
   1021 GGAGACG
```

App.11: The nucleotide sequence of *etfA*$_2$ cloned from *C. difficile* DSMZ 1296^T. The gray highlighted sequences refer to the *Esp3I* restriction site. The stop codon was replaced with *GGG*. There are two mutation highlighted with red.

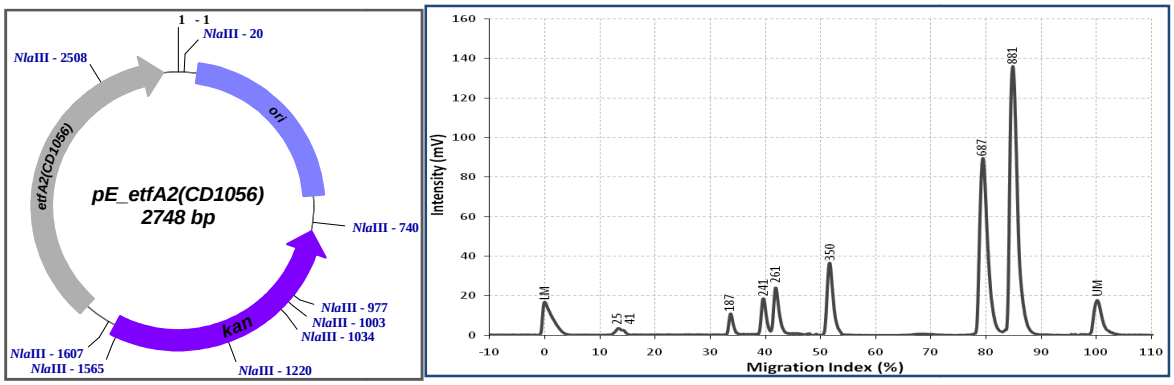
```
pE_CD1056-insert  CGTCTCCAAT GGGTAACGTT TTAGTAGTAA TAGAACAAAG AGAAAATGTA
    CD1056-etfA2  --------AT GGGTAACGTT TTAGTAGTAA TAGAACAAAG AGAAAATGTA

pE_CD1056-insert  ATTCAAACTG TTTCTTTAGA ATTACTAGGA AAGGCTACAG AAATAGCAAA
    CD1056-etfA2  ATTCAAACTG TTTCTTTAGA ATTACTAGGA AAGGCTACAG AAATAGCAAA

pE_CD1056-insert  AGATTATGAT ACAAAAGTTT CTGCATTACT TTTAGGTAGT AAGGTAGAAG
    CD1056-etfA2  AGATTATGAT ACAAAAGTTT CTGCATTACT TTTAGGTAGT AAGGTAGAAG

pE_CD1056-insert  GTTTAATAGA TACATTAGCA CACTATGGTG CAGATGAGGT AATAGTAGTA
    CD1056-etfA2  GTTTAATAGA TACATTAGCA CACTATGGTG CAGATGAGGT AATAGTAGTA

pE_CD1056-insert  GATGATGAAG CTTTAGCAGT GTATACAACT GAACCATATA CAAAAGCAGC
    CD1056-etfA2  GATGATGAAG CTTTAGCAGT GTATACAACT GAACCATATA CAAAAGCAGC

pE_CD1056-insert  TTATGAAGCA ATAAAAGCAG CTGACCCTAT AGTTGTATTA TTTGGTGCAA
    CD1056-etfA2  TTATGAAGCA ATAAAAGCAG CTGACCCTAT AGTTGTATTA TTTGGTGCAA

pE_CD1056-insert  CTTCAATAGG TAGAGATTTA GCGCCTAGAG TTTCTGCTAG AATACATACA
    CD1056-etfA2  CTTCAATAGG TAGAGATTTA GCTCCTAGAG TTTCTGCTAG AATACATACA

pE_CD1056-insert  GGTCTTACTG CTGACTGTAC AGGTCTTGCA GTAGCTGAAG ATACAAAATT
    CD1056-etfA2  GGTCTTACTG CTGACTGTAC AGGTCTTGCA GTAGCTGAAG ATACAAAATT

pE_CD1056-insert  ATTATTAATG ACAAGACCTG CCTTTGGTGG AAATATAATG GCAACAATAG
    CD1056-etfA2  ATTATTAATG ACAAGACCTG CATTTGGTGG AAATATAATG GCAACAATAG

pE_CD1056-insert  TTTGTAAAGA TTTCAGACCT CAAATGTCTA CAGTTAGACC AGGGGTTATG
    CD1056-etfA2  TTTGTAAAGA TTTCAGACCT CAAATGTCTA CAGTTAGACC AGGGGTTATG

pE_CD1056-insert  AAGAAAAATG AACCTGATGA AACTAAAGAA GCTGTAATTA ACCGTTTCAA
    CD1056-etfA2  AAGAAAAATG AACCTGATGA AACTAAAGAA GCTGTAATTA ACCGTTTCAA

pE_CD1056-insert  GGTAGAATTT AATGATGCTG ATAAATTAGT TCAAGTTGTA CAAGTAATAA
    CD1056-etfA2  GGTAGAATTT AATGATGCTG ATAAATTAGT TCAAGTTGTA CAAGTAATAA

pE_CD1056-insert  AAGAAGCTAA AAAACAAGTT AAAATAGAAG ATGCTAAGAT ATTAGTTTCT
    CD1056-etfA2  AAGAAGCTAA AAAACAAGTT AAAATAGAAG ATGCTAAGAT ATTAGTTTCT
```

```
pE_CD1056-insert   GCTGGACGTG GAATGGGTGG AAAAGAAAAC TTAGACATAC TTTATGAATT
    CD1056-etfA2   GCTGGACGTG GAATGGGTGG AAAAGAAAAC TTAGACATAC TTTATGAATT

pE_CD1056-insert   AGCTGAAATT ATAGGTGGAG AAGTTTCTGG TTCTCGTGCC ACTATAGATG
    CD1056-etfA2   AGCTGAAATT ATAGGTGGAG AAGTTTCTGG TTCTCGTGCC ACTATAGATG

pE_CD1056-insert   CAGGTTGGTT AGATAAAGCA AGACAAGTTG GTCAAACTGG TAAAACTGTA
    CD1056-etfA2   CAGGTTGGTT AGATAAAGCA AGACAAGTTG GTCAAACTGG TAAAACTGTA

pE_CD1056-insert   AGACCAGACC TTTATATAGC ATGTGGTATA TCTGGAGCAA TACAACATAT
    CD1056-etfA2   AGACCAGACC TTTATATAGC ATGTGGTATA TCTGGAGCAA TACAACATAT

pE_CD1056-insert   AGCTGGTATG GAAGATGCTG AGTTTATAGT TGCTATAAAT AAAAATCCAG
    CD1056-etfA2   AGCTGGTATG GAAGATGCTG AGTTTATAGT TGCTATAAAT AAAAATCCAG

pE_CD1056-insert   AAGCTCCAAT ATTTAAATAT GCTGATGTTG GTATAGTTGG AGATGTTCAT
    CD1056-etfA2   AAGCTCCAAT ATTTAAATAT GCTGATGTTG GTATAGTTGG AGATGTTCAT

pE_CD1056-insert   AAAGTGCTTC CAGAACTTAT CAGTCAGTTA AGTGTTGCAA AAGAAAAAGG
    CD1056-etfA2   AAAGTGCTTC CAGAACTTAT CAGTCAGTTA AGTGTTGCAA AAGAAAAAGG

pE_CD1056-insert   TGAAGTTTTA GCTAACGGGA GGAGACG
    CD1056-etfA2   TGAAGTTTTA GCTAACTAA- -------
```

App.12: Sequence alignment for the cloned *etfA*$_2$ from *C. difficile* DSMZ 1296^T and the genome derived genes from strain 630 [131] (GeneBank accession number CD1056). There are two mutations highlighted with red.

```
       pE_CD1056   MGNVLVVIEQ RENVIQTVSL ELLGKATEIA KDYDTKVSAL LLGSKVEGLI
           etfA2   MGNVLVVIEQ RENVIQTVSL ELLGKATEIA KDYDTKVSAL LLGSKVEGLI

       pE_CD1056   DTLAHYGADE VIVVDDEALA VYTTEPYTKA AYEAIKAADP IVVLFGATSI
           etfA2   DTLAHYGADE VIVVDDEALA VYTTEPYTKA AYEAIKAADP IVVLFGATSI

       pE_CD1056   GRDLAPRVSA RIHTGLTADC TGLAVAEDTK LLLMTRPAFG GNIMATIVCK
           etfA2   GRDLAPRVSA RIHTGLTADC TGLAVAEDTK LLLMTRPAFG GNIMATIVCK

       pE_CD1056   DFRPQMSTVR PGVMKKNEPD ETKEAVINRF KVEFNDADKL VQVVQVIKEA
           etfA2   DFRPQMSTVR PGVMKKNEPD ETKEAVINRF KVEFNDADKL VQVVQVIKEA

       pE_CD1056   KKQVKIEDAK ILVSAGRGMG GKENLDILYE LAEIIGGEVS GSRATIDAGW
           etfA2   KKQVKIEDAK ILVSAGRGMG GKENLDILYE LAEIIGGEVS GSRATIDAGW

       pE_CD1056   LDKARQVGQT GKTVRPDLYI ACGISGAIQH IAGMEDAEFI VAINKNPEAP
           etfA2   LDKARQVGQT GKTVRPDLYI ACGISGAIQH IAGMEDAEFI VAINKNPEAP

       pE_CD1056   IFKYADVGIV GDVHKVLPEL ISQLSVAKEK GEVLANG
           etfA2   IFKYADVGIV GDVHKVLPEL ISQLSVAKEK GEVLAN.
```

App.13: Amino acid alignment for the cloned *etfA*$_2$ from *C. difficile* DSMZ 1296^T and the genome derived genes from strain 630 [131]. No mutation was observed in the amino acid sequence.



Expected Fragments (bp)	26+32	42	186	237	260	345	720	901

App.14: Analysis of *pE_etfA$_2$ (CD1056)* with MultiNA after digestion with *NlaIII* restriction enzyme. On the right hand, there is electropherogram for the plasmid. On the left hand, there is pDRAW file shows the recognition sites *of NlaIII* as well as the position of the sequencing primers.

```
>pE-CD1057-insert
      1 CGTCTCCAAT GAGTACAAGT GATGTTAAAG TTTATGAGAA TGTAGCTGTT GAAGTAGATG
     61 GAAATATATG TACAGTGAAA ATGAATAGAC CTAAAGCCCT TAATGCAATA AATTCAAAGA
    121 CTTTAGAAGA ACTTTATGAA GTATTTGTAG ATATTAATAA TGATGAAACT ATTGATGTTG
    181 TAATATTGAC AGGGGAAGGA AAGGCATTTG TAGCTGGAGC AGATATTGCA TACATGAAAG
    241 ATTTAGATGC TGTAGCTGCT AAAGATTTTA GTATCTTAGG AGCAAAAGCT TTTGGAGAAA
    301 TAGAAAATAG TAAAAAAGTA GTGATAGCTG CTGTAAACGG ATTTGCTTTA GGTGGAGGAT
    361 GTGAACTTGC AATGGCATGT GATATAAGAA TTGCATCTGC TAAAGCTAAA TTTGGTCAGC
    421 CAGAAGTAAC TCTTGGAATA ACTCCAGGAT ATGGAGGAAC TCAAAGGCTT ACAAGATTGG
    481 TTGGAATGGC AAAAGCAAAA GAATTAATCT TTACAGGTCA AGTTATAAAA GCTGATGAAG
    541 CTGAAAAAAT AGGGCTAGTA AATAGAGTCG TTGAGCCAGA CATTTTAATA GAAGAAGTTG
    601 AGAAATTAGC TAAGATAATA GCTAAAAATG CTCAGCTTGC AGTTAGATAC TCTAAAGAAG
    661 CAATACAACT TGGTGCTCAA ACTGATATAA ATACTGGAAT AGATATAGAA TCTAATTTAT
    721 TTGGTCTTTG TTTTTCAACT AAAGACCAAA AAGAAGGAAT GTCAGCTTTC GTTGAAAAGA
    781 GAGAAGCTAA CTTTATAAAA GGGGGGAGGA GACG
```

App.15: The nucleotide sequence of *crt$_2$* cloned from *C. difficile* DSMZ 1296^T. The gray highlighted sequences refer to the *Esp3I* restriction site. The stop codon was replaced with *GGG*. There are two mutations highlighted with red.

```
pE-CD1057-insert  CGTCTCCAAT GAGTACAAGT GATGTTAAAG TTTATGAGAA TGTAGCTGTT
     CD1057-crt2  --------AT GAGTACAAGT GATGTTAAAG TTTATGAGAA TGTAGCTGTT

pE-CD1057-insert  GAAGTAGATG GAAATATATG TACAGTGAAA ATGAATAGAC CTAAAGCCCT
     CD1057-crt2  GAAGTAGATG GAAATATATG TACAGTGAAA ATGAATAGAC CTAAAGCCCT

pE-CD1057-insert  TAATGCAATA AATTCAAAGA CTTTAGAAGA ACTTTATGAA GTATTTGTAG
    CD1057-crt2   TAATGCAATA AATTCAAAGA CTTTAGAAGA ACTTTATGAA GTATTTGTAG

pE-CD1057-insert  ATATTAATAA TGATGAAACT ATTGATGTTG TAATATTGAC AGGGGAAGGA
     CD1057-crt2  ATATTAATAA TGATGAAACT ATTGATGTTG TAATATTGAC AGGGGAAGGA

pE-CD1057-insert  AAGGCATTTG TAGCTGGAGC AGATATTGCA TACATGAAAG ATTTAGATGC
     CD1057-crt2  AAGGCATTTG TAGCTGGAGC AGATATTGCA TACATGAAAG ATTTAGATGC

pE-CD1057-insert  TGTAGCTGCT AAAGATTTTA GTATCTTAGG AGCAAAAGCT TTTGGAGAAA
     CD1057-crt2  TGTAGCTGCT AAAGATTTTA GTATCTTAGG AGCAAAAGCT TTTGGAGAAA

pE-CD1057-insert  TAGAAAATAG TAAAAAAGTA GTGATAGCTG CTGTAAACGG ATTTGCTTTA
     CD1057-crt2  TAGAAAATAG TAAAAAAGTA GTGATAGCTG CTGTAAACGG ATTTGCTTTA
```

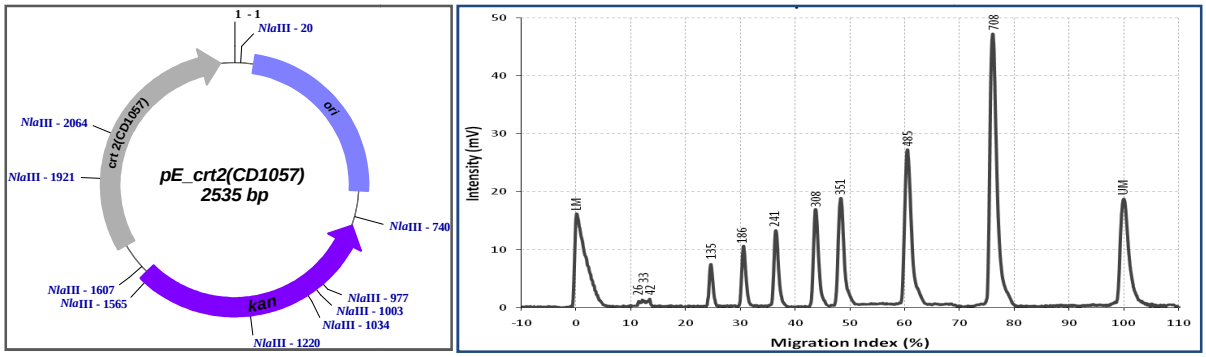
```
pE-CD1057-insert   GGTGGAGGAT GTGAACTTGC AATGGCATGT GATATAAGAA TTGCATCTGC
     CD1057-crt2   GGTGGAGGAT GTGAACTTGC AATGGCATGT GATATAAGAA TTGCATCTGC

pE-CD1057-insert   TAAAGCTAAA TTTGGTCAGC CAGAAGTAAC TCTTGGAATA ACTCCAGGAT
     CD1057-crt2   TAAAGCTAAA TTTGGTCAGC CAGAAGTAAC TCTTGGAATA ACTCCAGGAT

pE-CD1057-insert   ATGGAGGAAC TCAAAGGCTT ACAAGATTGG TTGGAATGGC AAAAGCAAAA
     CD1057-crt2   ATGGAGGAAC TCAAAGACTT ACAAGATTGG TTGGAATGGC AAAAGCAAAA

pE-CD1057-insert   GAATTAATCT TTACAGGTCA AGTTATAAAA GCTGATGAAG CTGAAAAAAT
     CD1057-crt2   GAATTAATCT TTACAGGTCA AGTTATAAAA GCTGATGAAG CTGAAAAAAT

pE-CD1057-insert   AGGGCTAGTA AATAGAGTCG TTGAGCCAGA CATTTTAATA GAAGAAGTTG
     CD1057-crt2   AGGGCTAATA AATAGAGTCG TTGAGCCAGA CATTTTAATA GAAGAAGTTG

pE-CD1057-insert   AGAAATTAGC TAAGATAATA GCTAAAAATG CTCAGCTTGC AGTTAGATAC
     CD1057-crt2   AGAAATTAGC TAAGATAATA GCTAAAAATG CTCAGCTTGC AGTTAGATAC

pE-CD1057-insert   TCTAAAGAAG CAATACAACT TGGTGCTCAA ACTGATATAA ATACTGGAAT
     CD1057-crt2   TCTAAAGAAG CAATACAACT TGGTGCTCAA ACTGATATAA ATACTGGAAT

pE-CD1057-insert   AGATATAGAA TCTAATTTAT TTGGTCTTTG TTTTTCAACT AAAGACCAAA
     CD1057-crt2   AGATATAGAA TCTAATTTAT TTGGTCTTTG TTTTTCAACT AAAGACCAAA

pE-CD1057-insert   AAGAAGGAAT GTCAGCTTTC GTTGAAAAGA GAGAAGCTAA CTTTATAAAA
     CD1057-crt2   AAGAAGGAAT GTCAGCTTTC GTTGAAAAGA GAGAAGCTAA CTTTATAAAA

pE-CD1057-insert   GGGGGGAGGA GACG
     CD1057-crt2   GGGTAA---- ----
```

App.16: Sequence alignment for the cloned *crt*$_2$ from *C. difficile* DSMZ 1296^T and the genome derived genes from strain 630 [131] (GeneBank accession number CD1057). There are two mutations highlighted with red.

```
       pE-CD1057   MSTSDVKVYE NVAVEVDGNI CTVKMNRPKA LNAINSKTLE ELYEVFVDIN
            Crt2   MSTSDVKVYE NVAVEVDGNI CTVKMNRPKA LNAINSKTLE ELYEVFVDIN

       pE-CD1057   NDETIDVVIL TGEGKAFVAG ADIAYMKDLD AVAAKDFSIL GAKAFGEIEN
            Crt2   NDETIDVVIL TGEGKAFVAG ADIAYMKDLD AVAAKDFSIL GAKAFGEIEN

       pE-CD1057   SKKVVIAAVN GFALGGGCEL AMACDIRIAS AKAKFGQPEV TLGITPGYGG
            Crt2   SKKVVIAAVN GFALGGGCEL AMACDIRIAS AKAKFGQPEV TLGITPGYGG

       pE-CD1057   TQRLTRLVGM AKAKELIFTG QVIKADEAEK IGLVNRVVEP DILIEEVEKL
            Crt2   TQRLTRLVGM AKAKELIFTG QVIKADEAEK IGLINRVVEP DILIEEVEKL

       pE-CD1057   AKIIAKNAQL AVRYSKEAIQ LGAQTDINTG IDIESNLFGL CFSTKDQKEG
            Crt2   AKIIAKNAQL AVRYSKEAIQ LGAQTDINTG IDIESNLFGL CFSTKDQKEG

       pE-CD1057   MSAFVEKREA NFIKGG
            Crt2   MSAFVEKREA NFIKG.
```

App.17: Amino acid alignment for the cloned *crt*$_2$ from *C. difficile* DSMZ 1296^T and the genome derived genes from strain 630 [131]. There is only one conservative amino acid exchanged, which provides a strain-to-strain deviation and is not PCR artefact. The stop codon was replaced with glycin

.



Expected Fragments (bp)	26	31	42	143	186	237	314	345	491	720

App.18: Analysis of *pE_crt$_2$ (CD1057)* with MultiNA after digestion with *NlaIII* restriction enzyme. On the right hand, there is electropherogram for the plasmid. On the left hand, there is pDRAW file shows the recognition sites of *NlaIII* as well as the position of the sequencing primers.

```
>pE_CD1058-insert
     1 CGTCTCCAAT GAAATTAGCT GTAATAGGTA GTGGAACTAT GGGAAGTGGT ATTGTACAAA
    61 CTTTTGCAAG TTGTGGACAT GATGTATGTT TAAAGAGTAG AACTCAAGGT GCTATAGATA
   121 AATGTTTAGC TTTATTAGAT AAAAATTTAA CTAAGTTAGT TACTAAGGGA AAAATGGATG
   181 AAGCTACAAA AGCAGAAATA TTAAGTCATG TTAGTTCAAC TACTAATTAT GAAGATTTAA
   241 AAGATATGGA TTTAATAATA GAAGCATCTG TAGAAGACAT GAATATAAAG AAAGATGTTT
   301 TCAAGTTACT AGATGAATTA TGTAAAGAAG ATACTATCTT GGCAACAAAT ACTTCATCAT
   361 TATCTATAAC AGAAATAGCT TCTTCTACTA AGAGACCAGA TAAAGTTATA GGAATGCATT
   421 TCTTTAATCC AGTTCCTATG ATGAAATTAG TTGAAGTTAT AAGTGGTCAG TTAACATCAA
   481 AAGTTACTTT TGATACAGTA TTTGAATTAT CTAAGAGTAT CAATAAAGTA CCAGTAGATG
   541 TATCTGAATC TCCTGGATTT GTAGTAAATA GAATACTTAT ACCTATGATA AATGAAGCTG
   601 TTGGTATATA TGCAGATGGT GTTGCAAGTA AAGAAGAAAT AGATGAAGCT ATGAAATTAG
   661 GAGCAAACCA TCCAATGGGA CCACTAGCAT TAGGTGATTT AATCGGATTA GATGTTGTTT
   721 TAGCTATAAT GAACGTTTTA TATACTGAAT TTGGAGATAC TAAATATAGA CCTCATCCAC
   781 TTTTAGCTAA AATGGTTAGA GCTAATCAAT TAGGAAGAAA AACTAAGATA GGATTCTATG
   841 ATTATAATAA AGGGAGGAGA CG
```

App.19: The nucleotide sequence of *hbd* cloned from *C. difficile* DSMZ 1296^T. The gray highlighted sequences refer to the *Esp3I* restriction site. The stop codon was replaced with *GGG*. There is only one mutation highlighted with red.

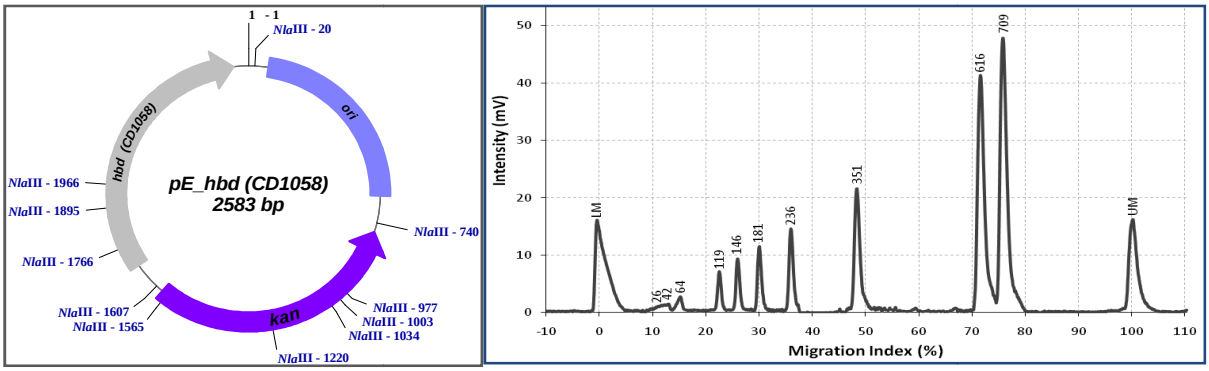
```
pE_CD1058-insert  CGTCTCCAAT GAAATTAGCT GTAATAGGTA GTGGAACTAT GGGAAGTGGT
      CD1058-hbd  --------AT GAAATTAGCT GTAATAGGTA GTGGAACTAT GGGAAGTGGT

pE_CD1058-insert  ATTGTACAAA CTTTTGCAAG TTGTGGACAT GATGTATGTT TAAAGAGTAG
      CD1058-hbd  ATTGTACAAA CTTTTGCAAG TTGTGGACAT GATGTATGTT TAAAGAGTAG

pE_CD1058-insert  AACTCAAGGT GCTATAGATA AATGTTTAGC TTTATTAGAT AAAAATTTAA
      CD1058-hbd  AACTCAAGGT GCTATAGATA AATGTTTAGC TTTATTAGAT AAAAATTTAA

pE_CD1058-insert  CTAAGTTAGT TACTAAGGGA AAAATGGATG AAGCTACAAA AGCAGAAATA
      CD1058-hbd  CTAAGTTAGT TACTAAGGGA AAAATGGATG AAGCTACAAA AGCAGAAATA

pE_CD1058-insert  TTAAGTCATG TTAGTTCAAC TACTAATTAT GAAGATTTAA AAGATATGGA
      CD1058-hbd  TTAAGTCATG TTAGTTCAAC TACTAATTAT GAAGATTTAA AAGATATGGA

pE_CD1058-insert  TTTAATAATA GAAGCATCTG TAGAAGACAT GAATATAAAG AAAGATGTTT
      CD1058-hbd  TTTAATAATA GAAGCATCTG TAGAAGACAT GAATATAAAG AAAGATGTTT

pE_CD1058-insert  TCAAGTTACT AGATGAATTA TGTAAAGAAG ATACTATCTT GGCAACAAAT
      CD1058-hbd  TCAAGTTACT AGATGAATTA TGTAAAGAAG ATACTATCTT GGCAACAAAT
```

```
pE_CD1058-insert   ACTTCATCAT TATCTATAAC AGAAATAGCT TCTTCTACTA AGAGACCAGA
      CD1058-hbd   ACTTCATCAT TATCTATAAC AGAAATAGCT TCTTCTACTA AGAGACCAGA

pE_CD1058-insert   TAAAGTTATA GGAATGCATT TCTTTAATCC AGTTCCTATG ATGAAATTAG
      CD1058-hbd   TAAAGTTATA GGAATGCATT TCTTTAATCC AGTTCCTATG ATGAAATTAG

pE_CD1058-insert   TTGAAGTTAT AAGTGGTCAG TTAACATCAA AAGTTACTTT TGATACAGTA
      CD1058-hbd   TTGAAGTTAT AAGTGGTCAG TTAACATCAA AAGTTACTTT TGATACAGTA

pE_CD1058-insert   TTTGAATTAT CTAAGAGTAT CAATAAAGTA CCAGTAGATG TATCTGAATC
      CD1058-hbd   TTTGAATTAT CTAAGAGTAT CAATAAAGTA CCAGTAGATG TATCTGAATC

pE_CD1058-insert   TCCTGGATTT GTAGTAAATA GAATACTTAT ACCTATGATA AATGAAGCTG
      CD1058-hbd   TCCTGGATTT GTAGTAAATA GAATACTTAT ACCTATGATA AATGAAGCTG

pE_CD1058-insert   TTGGTATATA TGCAGATGGT GTTGCAAGTA AAGAAGAAAT AGATGAAGCT
      CD1058-hbd   TTGGTATATA TGCAGATGGT GTTGCAAGTA AAGAAGAAAT AGATGAAGCT

pE_CD1058-insert   ATGAAATTAG GAGCAAACCA TCCAATGGGA CCACTAGCAT TAGGTGATTT
      CD1058-hbd   ATGAAATTAG GAGCAAACCA TCCAATGGGA CCACTAGCAT TAGGTGATTT

pE_CD1058-insert   AATCGGATTA GATGTTGTTT TAGCTATAAT GAACGTTTTA TATACTGAAT
      CD1058-hbd   AATTGGATTA GATGTTGTTT TAGCTATAAT GAACGTTTTA TATACTGAAT

pE_CD1058-insert   TTGGAGATAC TAAATATAGA CCTCATCCAC TTTTAGCTAA AATGGTTAGA
      CD1058-hbd   TTGGAGATAC TAAATATAGA CCTCATCCAC TTTTAGCTAA AATGGTTAGA

pE_CD1058-insert   GCTAATCAAT TAGGAAGAAA AACTAAGATA GGATTCTATG ATTATAATAA
      CD1058-hbd   GCTAATCAAT TAGGAAGAAA AACTAAGATA GGATTCTATG ATTATAATAA

pE_CD1058-insert   AGGGAGGAGA CG
      CD1058-hbd   ATAA------ --
```

App.20: Sequence alignment for the cloned *hbd* from *C. difficile* DSMZ 1296^T and the genome derived genes from strain 630 [131] (GeneBank accession number CD1058). There is only one mutation highlighted with red.

```
      pE_CD1058   MKLAVIGSGT MGSGIVQTFA SCGHDVCLKS RTQGAIDKCL ALLDKNLTKL
           BHBD   MKLAVIGSGT MGSGIVQTFA SCGHDVCLKS RTQGAIDKCL ALLDKNLTKL

      pE_CD1058   VTKGKMDEAT KAEILSHVSS TTNYEDLKDM DLIIEASVED MNIKKDVFKL
           BHBD   VTKGKMDEAT KAEILSHVSS TTNYEDLKDM DLIIEASVED MNIKKDVFKL

      pE_CD1058   LDELCKEDTI LATNTSSLSI TEIASSTKRP DKVIGMHFFN PVPMMKLVEV
           BHBD   LDELCKEDTI LATNTSSLSI TEIASSTKRP DKVIGMHFFN PVPMMKLVEV

      pE_CD1058   ISGQLTSKVT FDTVFELSKS INKVPVDVSE SPGFVVNRIL IPMINEAVGI
           BHBD   ISGQLTSKVT FDTVFELSKS INKVPVDVSE SPGFVVNRIL IPMINEAVGI

      pE_CD1058   YADGVASKEE IDEAMKLGAN HPMGPLALGD LIGLDVVLAI MNVLYTEFGD
           BHBD   YADGVASKEE IDEAMKLGAN HPMGPLALGD LIGLDVVLAI MNVLYTEFGD

      pE_CD1058   TKYRPHPLLA KMVRANQLGR KTKIGFYDYN KG
           BHBD   TKYRPHPLLA KMVRANQLGR KTKIGFYDYN K.
```

App.21: Amino acid alignment for the cloned *hbd* from *C. difficile* DSMZ 1296^T and the genome derived genes from strain 630 [131]. No mutation was observed in the amino acid sequence. The stop codon was replaced with glycin.



Expected fragments (bp)	26+31	42	71	129	159	186	237	345	637	720

App.22: Analysis of *pE_hbd (CD1058)* with MultiNA after digestion with *NlaIII* restriction enzyme. On the right hand, there is electropherogram for the plasmid. On the left hand, there is pDRAW file shows the recognition sites of *NlaIII* as well as the position of the sequencing primers.

```
>pE_CD1059-insert
      1 CGTCTCCAAT GAGAGAAGTA GTAATTGCCA GTGCAGCTAG AACAGCAGTA GGAAGTTTTG
     61 GAGGAGCATT TAAATCAGTT TCAGCGGTAG AGTTAGGGGT AACAGCAGCT AAAGAAGCTA
    121 TAAAAAGAGC TAACATAACT CCAGATATGA TAGATGAATC TCTTTTAGGG GGAGTACTTA
    181 CAGCAGGTCT TGGACAAAAT ATAGCAAGAC AAATAGCATT AGGAGCAGGA ATACCAGTAG
    241 AAAAACCAGC TATGACTATA AATATAGTTT GTGGTTCTGG ATTAAGATCT GTTTCAATGG
    301 CATCTCAACT TATAGCATTA GGTGATGCTG ATATAATGTT AGTTGGTGGA GCTGAAAACA
    361 TGAGTATGTC TCCTTATTTA GTACCAAGTG CGAGATATGG TGCAAGAATG GGTGATGCTG
    421 CTTTTGTTGA TTCAATGATA AAAGATGGAT TATCAGACAT ATTTAATAAC TATCACATGG
    481 GTATTACTGC TGAAAACATA GCAGAGCAAT GGAATATAAC TAGAGAAGAA CAAGATGAAT
    541 TAGCTCTTGC AAGTCAAAAT AAAGCTGAAA AAGCTCAAGC TGAAGGAAAA TTTGATGAAG
    601 AAATAGTTCC TGTTGTTATA AAAGGAAGAA AAGGTGACAC TGTAGTAGAT AAAGATGAAT
    661 ATATTAAGCC TGGCACTACA ATGGAGAAAC TTGCTAAGTT AAGACCTGCA TTTAAAAAAG
    721 ATGGAACAGT TACTGCTGGT AATGCATCAG GAATAAATGA TGGTGCTGCT ATGTTAGTAG
    781 TAATGGCTAA AGAAAAAGCT GAAGAACTAG GAATAGAGCC TCTTGCAACT ATAGTTTCTT
    841 ATGGAACAGC TGGTGTTGAC CCTAAAATAA TGGGATATGG ACCAGTTCCA GCAACTAAAA
    901 AAGCTTTAGA AGCTGCTAAT ATGACTATTG AAGATATAGA TTTAGTTGAA GCTAATGAGG
    961 CATTTGCTGC CCAATCTGTA GCTGTAATAA GAGACTTAAA TATAGATATG AATAAAGTTA
   1021 ATGTTAATGG TGGAGCAATA GCTATAGGAC ATCCAATAGG ATGCTCAGGA GCAAGAATAC
   1081 TTACTACACT TTTATATGAA ATGAAGAGAA GAGATGCTAA AACTGGTCTT GCTACACTTT
   1141 GTATAGGCGG TGGAATGGGA ACTACTTTAA TAGTTAAGAG AGGGAGGAGA CG
```

App.23: The nucleotide sequence of *thiA1* cloned from *C. difficile* DSMZ 1296^{T}. The gray highlighted sequences refer to the *Esp3I* restriction site. The stop codon was replaced with *GGG*. There are two mutations highlighted with red.

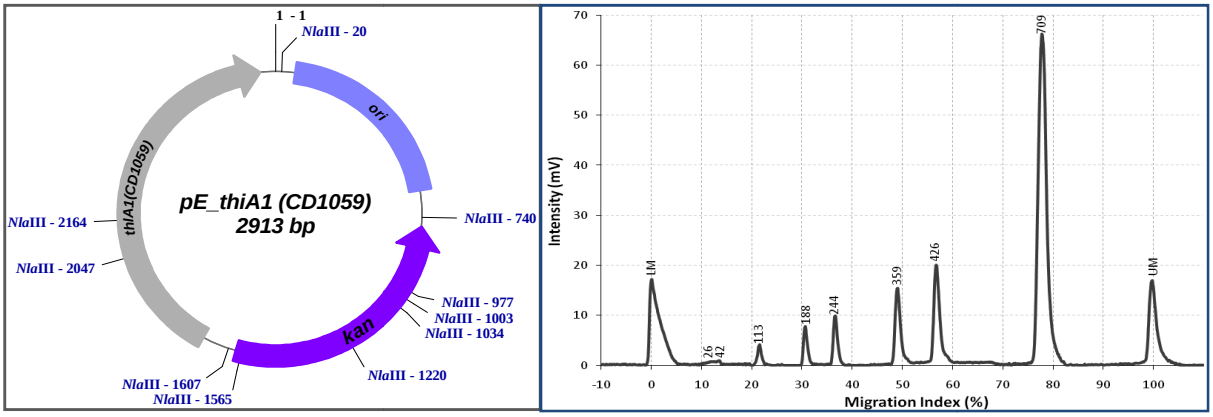
```
pE_CD1059-insert  CGTCTCCAAT GAGAGAAGTA GTAATTGCCA GTGCAGCTAG AACAGCAGTA
    CD1059-thiAI  --------AT GAGAGAAGTA GTAATTGCCA GTGCAGCTAG AACAGCAGTA

pE_CD1059-insert  GGAAGTTTTG GAGGAGCATT TAAATCAGTT TCAGCGGTAG AGTTAGGGGT
    CD1059-thiAI  GGAAGTTTTG GAGGAGCATT TAAATCAGTT TCAGCGGTAG AGTTAGGGGT

pE_CD1059-insert  AACAGCAGCT AAAGAAGCTA TAAAAAGAGC TAACATAACT CCAGATATGA
    CD1059-thiAI  AACAGCAGCT AAAGAAGCTA TAAAAAGAGC TAACATAACT CCAGATATGA

pE_CD1059-insert  TAGATGAATC TCTTTTAGGG GGAGTACTTA CAGCAGGTCT TGGACAAAAT
    CD1059-thiAI  TAGATGAATC TCTTTTAGGG GGAGTACTTA CAGCAGGTCT TGGACAAAAT

pE_CD1059-insert  ATAGCAAGAC AAATAGCATT AGGAGCAGGA ATACCAGTAG AAAAACCAGC
    CD1059-thiAI  ATAGCAAGAC AAATAGCATT AGGAGCAGGA ATACCAGTAG AAAAACCAGC
```

```
pE_CD1059-insert   TATGACTATA AATATAGTTT GTGGTTCTGG ATTAAGATCT GTTTCAATGG
    CD1059-thiAI   TATGACTATA AATATAGTTT GTGGTTCTGG ATTAAGATCT GTTTCAATGG

pE_CD1059-insert   CATCTCAACT TATAGCATTA GGTGATGCTG ATATAATGTT AGTTGGTGGA
    CD1059-thiAI   CATCTCAACT TATAGCATTA GGTGATGCTG ATATAATGTT AGTTGGTGGA

pE_CD1059-insert   GCTGAAAACA TGAGTATGTC TCCTTATTTA GTACCAAGTG CGAGATATGG
    CD1059-thiAI   GCTGAAAACA TGAGTATGTC TCCTTATTTA GTACCAAGTG CGAGATATGG

pE_CD1059-insert   TGCAAGAATG GGTGATGCTG CTTTTGTTGA TTCAATGATA AAAGATGGAT
    CD1059-thiAI   TGCAAGAATG GGTGATGCTG CTTTTGTTGA TTCAATGATA AAAGATGGAT

pE_CD1059-insert   TATCAGACAT ATTTAATAAC TATCACATGG GTATTACTGC TGAAAACATA
    CD1059-thiAI   TATCAGACAT ATTTAATAAC TATCACATGG GTATTACTGC TGAAAACATA

pE_CD1059-insert   GCAGAGCAAT GGAATATAAC TAGAGAAGAA CAAGATGAAT TAGCTCTTGC
    CD1059-thiAI   GCAGAGCAAT GGAATATAAC TAGAGAAGAA CAAGATGAAT TAGCTCTTGC

pE_CD1059-insert   AAGTCAAAAT AAAGCTGAAA AAGCTCAAGC TGAAGGAAAA TTTGATGAAG
    CD1059-thiAI   AAGTCAAAAT AAAGCTGAAA AAGCTCAAGC TGAAGGAAAA TTTGATGAAG

pE_CD1059-insert   AAATAGTTCC TGTTGTTATA AAAGGAAGAA AAGGTGACAC TGTAGTAGAT
    CD1059-thiAI   AAATAGTTCC TGTTGTTATA AAAGGAAGAA AAGGTGACAC TGTAGTAGAT

pE_CD1059-insert   AAAGATGAAT ATATTAAGCC TGGCACTACA ATGGAGAAAC TTGCTAAGTT
    CD1059-thiAI   AAAGATGAAT ATATTAAGCC TGGCACTACA ATGGAGAAAC TTGCTAAGTT

pE_CD1059-insert   AAGACCTGCA TTTAAAAAAG ATGGAACAGT TACTGCTGGT AATGCATCAG
    CD1059-thiAI   AAGACCTGCA TTTAAAAAAG ATGGAACAGT TACTGCTGGT AATGCATCAG

pE_CD1059-insert   GAATAAATGA TGGTGCTGCT ATGTTAGTAG TAATGGCTAA AGAAAAAGCT
    CD1059-thiAI   GAATAAATGA TGGTGCTGCT ATGCTAGTAG TAATGGCTAA AGAAAAAGCT

pE_CD1059-insert   GAAGAACTAG GAATAGAGCC TCTTGCAACT ATAGTTTCTT ATGGAACAGC
    CD1059-thiAI   GAAGAACTAG GAATAGAGCC TCTTGCAACT ATAGTTTCTT ATGGAACAGC

pE_CD1059-insert   TGGTGTTGAC CCTAAAATAA TGGGATATGG ACCAGTTCCA GCAACTAAAA
    CD1059-thiAI   TGGTGTTGAC CCTAAAATAA TGGGATATGG ACCAGTTCCA GCAACTAAAA

pE_CD1059-insert   AAGCTTTAGA AGCTGCTAAT ATGACTATTG AAGATATAGA TTTAGTTGAA
    CD1059-thiAI   AAGCTTTAGA AGCTGCTAAT ATGACTATTG AAGATATAGA TTTAGTTGAA

pE_CD1059-insert   GCTAATGAGG CATTTGCTGC CCAATCTGTA GCTGTAATAA GAGACTTAAA
    CD1059-thiAI   GCTAATGAGG CATTTGCTGC CCAATCTGTA GCTGTAATAA GAGACTTAAA

pE_CD1059-insert   TATAGATATG AATAAAGTTA ATGTTAATGG TGGAGCAATA GCTATAGGAC
    CD1059-thiAI   TATAGATATG AATAAAGTTA ATGTTAATGG TGGAGCAATA GCTATAGGAC

pE_CD1059-insert   ATCCAATAGG ATGCTCAGGA GCAAGAATAC TTACTACACT TTTATATGAA
    CD1059-thiAI   ATCCAATAGG ATGCTCAGGA GCAAGAATAC TTACTACACT TTTATATGAA

pE_CD1059-insert   ATGAAGAGAA GAGATGCTAA AACTGGTCTT GCTACACTTT GTATAGGCGG
    CD1059-thiAI   ATGAAGAGAA GAGATGCTAA AACTGGTCTT GCTACACTTT GTATAGGTGG

pE_CD1059-insert   TGGAATGGGA ACTACTTTAA TAGTTAAGAG AGGGAGGAGA CG
    CD1059-thiAI   TGGAATGGGA ACTACTTTAA TAGTTAAGAG ATAG------ --
```

App.24: Sequence alignment for the cloned *thiA1* from *C. difficile* DSMZ 1296^{T} and the genome derived genes from strain 630 [131] (GeneBank accession number CD1059). There are two mutations highlighted with red.

```
pE_CD1059  MREVVIASAA RTAVGSFGGA FKSVSAVELG VTAAKEAIKR ANITPDMIDE
    ThiAI  MREVVIASAA RTAVGSFGGA FKSVSAVELG VTAAKEAIKR ANITPDMIDE

pE_CD1059  SLLGGVLTAG LGQNIARQIA LGAGIPVEKP AMTINIVCGS GLRSVSMASQ
    ThiAI  SLLGGVLTAG LGQNIARQIA LGAGIPVEKP AMTINIVCGS GLRSVSMASQ

pE_CD1059  LIALGDADIM LVGGAENMSM SPYLVPSARY GARMGDAAFV DSMIKDGLSD
    ThiAI  LIALGDADIM LVGGAENMSM SPYLVPSARY GARMGDAAFV DSMIKDGLSD

pE_CD1059  IFNNYHMGIT AENIAEQWNI TREEQDELAL ASQNKAEKAQ AEGKFDEEIV
    ThiAI  IFNNYHMGIT AENIAEQWNI TREEQDELAL ASQNKAEKAQ AEGKFDEEIV

pE_CD1059  PVVIKGRKGD TVVDKDEYIK PGTTMEKLAK LRPAFKKDGT VTAGNASGIN
    ThiAI  PVVIKGRKGD TVVDKDEYIK PGTTMEKLAK LRPAFKKDGT VTAGNASGIN

pE_CD1059  DGAAMLVVMA KEKAEELGIE PLATIVSYGT AGVDPKIMGY GPVPATKKAL
    ThiAI  DGAAMLVVMA KEKAEELGIE PLATIVSYGT AGVDPKIMGY GPVPATKKAL

pE_CD1059  EAANMTIEDI DLVEANEAFA AQSVAVIRDL NIDMNKVNVN GGAIAIGHPI
    ThiAI  EAANMTIEDI DLVEANEAFA AQSVAVIRDL NIDMNKVNVN GGAIAIGHPI

pE_CD1059  GCSGARILTT LLYEMKRRDA KTGLATLCIG GGMGTTLIVK RG
    ThiAI  GCSGARILTT LLYEMKRRDA KTGLATLCIG GGMGTTLIVK R.
```

App.25: Amino acid alignment for the cloned *thiA1* from *C. difficile* DSMZ 1296^T and the genome derived genes from strain 630 [131]. No mutation was observed in the amino acid sequence. The stop codon was replaced with glycin.



Expected fragments (bp)	26+31	42	117	186	237	345	440	720+769

App.26: Analysis of *pE_thiA1 (CD1059)* with MultiNA after digestion with *NlaIII* restriction enzyme. On the right hand, there is electropherogram for the plasmid. On the left hand, there is pDRAW file shows the recognition sites of *NlaIII* as well as the position of the sequencing primers.

```
>pE-CD0112-insert
      1 CGTCTCCAAT GAGAAGTTTT GAAGAAGTAA TTAAGTTTGC AAAAGAAAGA GGACCTAAAA
     61 CTATATCAGT AGCATGTTGC CAAGATAAAG AAGTTTTAAT GGCAGTTGAA ATGGCTAGAA
    121 AAGAAAAAAT AGCAAATGCC ATTTTAGTAG GAGATATAGA AAAGACTAAA GAAATTGCAA
    181 AAAGCATAGA CATGGATATC GAAAATTATG AACTGATAGA TATAAAAGAT TTAGCAGAAG
    241 CATCTCTAAA ATCTGTTGAA TTAGTTTCAC AAGGAAAAGC CGACATGGTA ATGAAAGGCT
    301 TAGTAGACAC ATCAATAATA CTAAAAGCAG TTTTAAATAA AGAAGTAGGT CTTAGAACTG
    361 GAAATGTATT AAGTCACGTA GCAGTATTTG ATGTAGAGGG ATATGATAGA TTATTTTTCG
    421 TAACTGACGC AGCTATGAAC TTAGCTCCTG ATACAAATAC TAAAAAGCAA ATCATAGAAA
    481 ATGCTTGCAC AGTAGCACAT TCATTAGATA TAAGTGAACC AAAAGTTGCT GCAATATGCG
    541 CAAAAGAAAA AGTAAATCCA AAAATGAAAG ATACAGTTGA AGCTAAAGAA CTAGAAGAAA
    601 TGTATGAAAG AGGAGAAATC AAAGGTTGTA TGGTTGGTGG GCCTTTTGCA ATTGATAATG
    661 CAGTATCTTT AGAAGCAGCT AAACATAAAG GTATAAATCA TCCTGTAGCA GGACGAGCTG
    721 ATATATTATT AGCCCCAGAT ATTGAAGGTG GTAACATATT ATATAAAGCT TTGGTATTCT
    781 TCTCAAAATC AAAAAATGCA GGAGTTATAG TTGGGGCTAA AGCACCAATA ATATTAACTT
    841 CTAGAGCAGA CAGTGAAGAA ACTAAACTAA ACTCAATAGC TTTAGGTGTT TTAATGGCAG
    901 CAAAGGCAGG GAGGAGACG
```

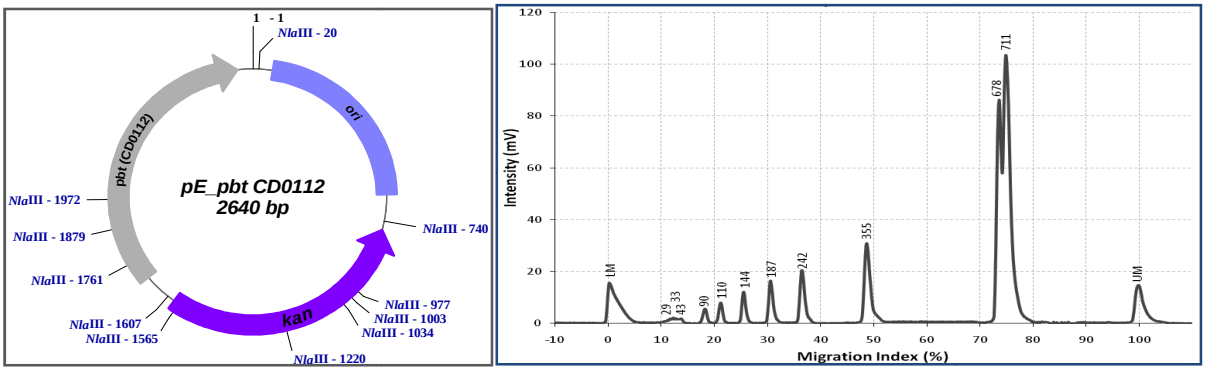
App.27: The nucleotide sequence of *pbt* cloned from *C. difficile* DSMZ 1296^T. The gray highlighted sequences refer to the *Esp3I* restriction site. There are two mutations, one of which was introduced on purpose to remove a *LguI*-recognition site and it is green highlighted. The second one observed is a strain-to-strain difference and highlighted with red.

```
pE-CD0112-insert  CGTCTCCAAT GAGAAGTTTT GAAGAAGTAA TTAAGTTTGC AAAAGAAAGA
      CD0112-pbt  --------AT GAGAAGTTTT GAAGAAGTAA TTAAGTTTGC AAAAGAAAGA

pE-CD0112-insert  GGACCTAAAA CTATATCAGT AGCATGTTGC CAAGATAAAG AAGTTTTAAT
      CD0112-pbt  GGACCTAAAA CTATATCAGT AGCATGTTGC CAAGATAAAG AAGTTTTAAT

pE-CD0112-insert  GGCAGTTGAA ATGGCTAGAA AAGAAAAAAT AGCAAATGCC ATTTTAGTAG
      CD0112-pbt  GGCAGTTGAA ATGGCTAGAA AAGAAAAAAT AGCAAATGCC ATTTTAGTAG

pE-CD0112-insert  GAGATATAGA AAAGACTAAA GAAATTGCAA AAAGCATAGA CATGGATATC
      CD0112-pbt  GAGATATAGA AAAGACTAAA GAAATTGCAA AAAGCATAGA CATGGATATC

pE-CD0112-insert  GAAAATTATG AACTGATAGA TATAAAAGAT TTAGCAGAAG CATCTCTAAA
      CD0112-pbt  GAAAATTATG AACTGATAGA TATAAAAGAT TTAGCAGAAG CATCTCTAAA

pE-CD0112-insert  ATCTGTTGAA TTAGTTTCAC AAGGAAAAGC CGACATGGTA ATGAAAGGCT
      CD0112-pbt  ATCTGTTGAA TTAGTTTCAC AAGGAAAAGC CGACATGGTA ATGAAAGGCT

pE-CD0112-insert  TAGTAGACAC ATCAATAATA CTAAAAGCAG TTTTAAATAA AGAAGTAGGT
      CD0112-pbt  TAGTAGACAC ATCAATAATA CTAAAAGCAG TTTTAAATAA AGAAGTAGGT

pE-CD0112-insert  CTTAGAACTG GAAATGTATT AAGTCACGTA GCAGTATTTG ATGTAGAGGG
      CD0112-pbt  CTTAGAACTG GAAATGTATT AAGTCACGTA GCAGTATTTG ATGTAGAGGG

pE-CD0112-insert  ATATGATAGA TTATTTTTCG TAACTGACGC AGCTATGAAC TTAGCTCCTG
      CD0112-pbt  ATATGATAGA TTATTTTTCG TAACTGACGC AGCTATGAAC TTAGCTCCTG

pE-CD0112-insert  ATACAAATAC TAAAAAGCAA ATCATAGAAA ATGCTTGCAC AGTAGCACAT
      CD0112-pbt  ATACAAATAC TAAAAAGCAA ATCATAGAAA ATGCTTGCAC AGTAGCACAT

pE-CD0112-insert  TCATTAGATA TAAGTGAACC AAAAGTTGCT GCAATATGCG CAAAAGAAAA
      CD0112-pbt  TCATTAGATA TAAGTGAACC AAAAGTTGCT GCAATATGCG CAAAAGAAAA

pE-CD0112-insert  AGTAAATCCA AAAATGAAAG ATACAGTTGA AGCTAAAGAA CTAGAAGAAA
      CD0112-pbt  AGTAAATCCA AAAATGAAAG ATACAGTTGA AGCTAAAGAA CTAGAAGAAA

pE-CD0112-insert  TGTATGAAAG AGGAGAAATC AAAGGTTGTA TGGTTGGTGG GCCTTTTGCA
      CD0112-pbt  TGTATGAAAG AGGAGAAATC AAAGGTTGTA TGGTTGGTGG ACCTTTTGCA

pE-CD0112-insert  ATTGATAATG CAGTATCTTT AGAAGCAGCT AAACATAAAG GTATAAATCA
      CD0112-pbt  ATTGATAATG CAGTATCTTT AGAAGCAGCT AAACATAAAG GTATAAATCA

pE-CD0112-insert  TCCTGTAGCA GGACGAGCTG ATATATTATT AGCCCCAGAT ATTGAAGGTG
      CD0112-pbt  TCCTGTAGCA GGAAGAGCTG ATATATTATT AGCCCCAGAT ATTGAAGGTG

pE-CD0112-insert  GTAACATATT ATATAAAGCT TTGGTATTCT TCTCAAAATC AAAAAATGCA
      CD0112-pbt  GTAACATATT ATATAAAGCT TTGGTATTCT TCTCAAAATC AAAAAATGCA

pE-CD0112-insert  GGAGTTATAG TTGGGGCTAA AGCACCAATA ATATTAACTT CTAGAGCAGA
      CD0112-pbt  GGAGTTATAG TTGGGGCTAA AGCACCAATA ATATTAACTT CTAGAGCAGA

pE-CD0112-insert  CAGTGAAGAA ACTAAACTAA ACTCAATAGC TTTAGGTGTT TTAATGGCAG
      CD0112-pbt  CAGTGAAGAA ACTAAACTAA ACTCAATAGC TTTAGGTGTT TTAATGGCAG

pE-CD0112-insert  CAAAGGCAGG GAGGAGACG
      CD0112-pbt  CAAAGGCATA A--------
```

App.28: Sequence alignment for the cloned *pbt* from *C. difficile* DSMZ 1296^T and the genome derived genes from strain 630 [131] (GeneBank accession number CD0112). There are two mutations, one of which was introduced on purpose to remove a *LguI*-recognition site and it is green highlighted. The second one observed is a strain-to-strain difference and highlighted with red.

```
pE-CD0112   MRSFEEVIKF AKERGPKTIS VACCQDKEVL MAVEMARKEK IANAILVGDI
      Pbt   MRSFEEVIKF AKERGPKTIS VACCQDKEVL MAVEMARKEK IANAILVGDI

pE-CD0112   EKTKEIAKSI DMDIENYELI DIKDLAEASL KSVELVSQGK ADMVMKGLVD
      Pbt   EKTKEIAKSI DMDIENYELI DIKDLAEASL KSVELVSQGK ADMVMKGLVD

pE-CD0112   TSIILKAVLN KEVGLRTGNV LSHVAVFDVE GYDRLFFVTD AAMNLAPDTN
      Pbt   TSIILKAVLN KEVGLRTGNV LSHVAVFDVE GYDRLFFVTD AAMNLAPDTN

pE-CD0112   TKKQIIENAC TVAHSLDISE PKVAAICAKE KVNPKMKDTV EAKELEEMYE
      Pbt   TKKQIIENAC TVAHSLDISE PKVAAICAKE KVNPKMKDTV EAKELEEMYE

pE-CD0112   RGEIKGCMVG GPFAIDNAVS LEAAKHKGIN HPVAGRADIL LAPDIEGGNI
      Pbt   RGEIKGCMVG GPFAIDNAVS LEAAKHKGIN HPVAGRADIL LAPDIEGGNI

pE-CD0112   LYKALVFFSK SKNAGVIVGA KAPIILTSRA DSEETKLNSI ALGVLMAAKA
      Pbt   LYKALVFFSK SKNAGVIVGA KAPIILTSRA DSEETKLNSI ALGVLMAAKA

pE-CD0112   G
      Pbt   .
```

App.29: Amino acid alignment for the cloned *pbt* from *C. difficile* DSMZ 1296^{T} and the genome derived genes from strain 630 [131]. No mutation was observed in the amino acid sequence. The stop codon was replaced with glycin.



Expected fragments (bp)	26	31	42	93	118	154	186	237	345	688	720

App.30: Analysis of *pE_pbt (CD0112)* with MultiNA after digestion with *NlaIII* restriction enzyme. On the right hand, there is electropherogram for the plasmid. On the left hand, there is pDRAW file shows the recognition sites of *NlaIII* as well as the sequencing primers positions.

```
pE_CD0113-insert
     1 CGTCTCCAAT GAGCAAAATA TTTAAAATCT TAACAATAAA TCCTGGTTCG ACATCAACTA
    61 AAATAGCTGT ATTTGATAAT GAGGATTTAG TATTTGAAAA AACTTTAAGA CATTCTTCAG
   121 AAGAAATAGG AAAATATGAG AAGGTGTCTG ACCAATTTGA ATTTCGTAAA CAAGTAATAG
   181 AAGAAGCTCT AAAAGAAGGT GGAGTAAAAA CATCTGAATT AGATGCTGTA GTAGGTAGAG
   241 GAGGACTTCT TAAACCTATA AAAGGTGGTA CTTATTCAGT AAGTGCTGCT ATGATTGAAG
   301 ATTTAAAAGT GGGAGTTTTA GGAGAACACG CTTCAAACCT AGGTGGAATA ATAGCAAAAC
   361 AAATAGGTGA AGAAGTAAAT GTTCCTTCAT ACATAGTAGA CCCTGTTGTT GTAGATGAAT
   421 TAGAAGATGT TGCTAGAATT TCTGGTATGC CTGAAATAAG TAGAGCAAGT GTAGTACATG
   481 CTTTAAATCA AAAGGCAATA GCAAGAAGAT ATGCTAGAGA AATAAACAAG AAATATGAAG
   541 ATATAAATCT TATAGTTGCA CACATGGGTG GAGGAGTTTC TGTTGGAGCT CATAAAAATG
   601 GTAAAATAGT AGATGTTGCA AACGCATTAG ATGGAGAAGG ACCTTTCTCT CCAGAAAGAA
   661 GTGGTGGACT ACCAGTAGGT GCATTAGTAA AAATGTGCTT TAGTGGAAAA TATACTCAAG
   721 ATGAAATTAA AAAGAAAATA AAAGGTAATG GCGGACTAGT TGCATACTTA AACACTAATG
   781 ATGCTAGAGA AGTTGAAGAA AGAATTGAAG CTGGTGATGA AAAAGCTAAA TTAGTATATG
```

```
   841 AAGCTATGGC ATATCAAATC TCTAAAGAAA TAGGAGCTAG TGCTGCAGTT CTTAAGGGAG
   901 ATGTAAAAGC AATATTATTA ACTGGTGGAA TCGCATATTC AAAAATGTTT ACAGAAATGA
   961 TTGCAGATAG AGTTAAATTT ATAGCAGATG TAAAAGTTTA TCCAGGTGAA GATGAAATGA
  1021 TTGCATTAGC TCAAGGTGGA CTTAGAGTTT TAACTGGTGA AGAAGAGGCT CAAGTTTATG
  1081 ATAACGGGAG GAGACG
```

App.31: The nucleotide sequence of *buk* cloned from *C. difficile* DSMZ 1296^{T}. The gray highlighted sequences refer to the *Esp3I* restriction site. There are two mutations highlighted with red.

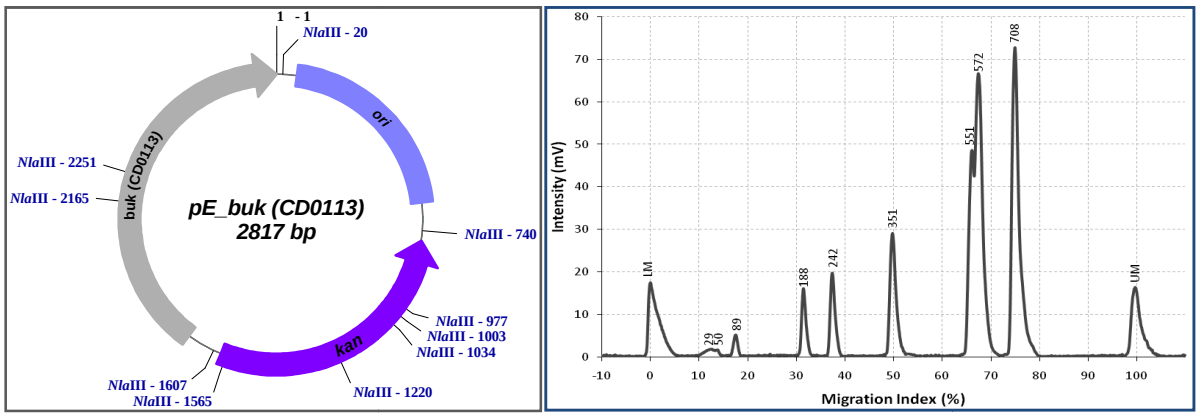
```
pE_CD0113-insert  CGTCTCCAAT GAGCAAAATA TTTAAAATCT TAACAATAAA TCCTGGTTCG
      CD0113-buk  --------AT GAGCAAAATA TTTAAAATCT TAACAATAAA TCCTGGTTCG

pE_CD0113-insert  ACATCAACTA AAATAGCTGT ATTTGATAAT GAGGATTTAG TATTTGAAAA
      CD0113-buk  ACATCAACTA AAATAGCTGT ATTTGATAAT GAGGATTTAG TATTTGAAAA

pE_CD0113-insert  AACTTTAAGA CATTCTTCAG AAGAAATAGG AAAATATGAG AAGGTGTCTG
      CD0113-buk  AACTTTAAGA CATTCTTCAG AAGAAATAGG AAAATATGAG AAGGTGTCTG

pE_CD0113-insert  ACCAATTTGA ATTTCGTAAA CAAGTAATAG AAGAAGCTCT AAAAGAAGGT
      CD0113-buk  ACCAATTTGA ATTTCGTAAA CAAGTAATAG AAGAAGCTCT AAAAGAAGGT

pE_CD0113-insert  GGAGTAAAAA CATCTGAATT AGATGCTGTA GTAGGTAGAG GAGGACTTCT
      CD0113-buk  GGAGTAAAAA CATCTGAATT AGATGCTGTA GTAGGTAGAG GAGGACTTCT

pE_CD0113-insert  TAAACCTATA AAAGGTGGTA CTTATTCAGT AAGTGCTGCT ATGATTGAAG
      CD0113-buk  TAAACCTATA AAAGGTGGTA CTTATTCAGT AAGTGCTGCT ATGATTGAAG

pE_CD0113-insert  ATTTAAAAGT GGGAGTTTTA GGAGAACACG CTTCAAACCT AGGTGGAATA
      CD0113-buk  ATTTAAAAGT GGGAGTTTTA GGAGAACACG CTTCAAACCT AGGTGGAATA

pE_CD0113-insert  ATAGCAAAAC AAATAGGTGA AGAAGTAAAT GTTCCTTCAT ACATAGTAGA
      CD0113-buk  ATAGCAAAAC AAATAGGTGA AGAAGTAAAT GTTCCTTCAT ACATAGTAGA

pE_CD0113-insert  CCCTGTTGTT GTAGATGAAT TAGAAGATGT TGCTAGAATT TCTGGTATGC
      CD0113-buk  CCCTGTTGTT GTAGATGAAT TAGAAGATGT TGCTAGAATT TCTGGTATGC

pE_CD0113-insert  CTGAAATAAG TAGAGCAAGT GTAGTACATG CTTTAAATCA AAAGGCAATA
      CD0113-buk  CTGAAATAAG TAGAGCAAGT GTAGTACATG CTTTAAATCA AAAGGCAATA

pE_CD0113-insert  GCAAGAAGAT ATGCTAGAGA AATAAACAAG AAATATGAAG ATATAAATCT
      CD0113-buk  GCAAGAAGAT ATGCTAGAGA AATAAACAAG AAATATGAAG ATATAAATCT

pE_CD0113-insert  TATAGTTGCA CACATGGGTG GAGGAGTTTC TGTTGGAGCT CATAAAAATG
      CD0113-buk  TATAGTTGCA CATATGGGTG GAGGAGTTTC TGTTGGAGCT CATAAAAATG

pE_CD0113-insert  GTAAAATAGT AGATGTTGCA AACGCATTAG ATGGAGAAGG ACCTTTCTCT
      CD0113-buk  GTAAAATAGT AGATGTTGCA AACGCATTAG ATGGAGAAGG ACCTTTCTCT

pE_CD0113-insert  CCAGAAAGAA GTGGTGGACT ACCAGTAGGT GCATTAGTAA AAATGTGCTT
      CD0113-buk  CCAGAAAGAA GTGGTGGACT ACCAGTAGGT GCATTAGTAA AAATGTGCTT

pE_CD0113-insert  TAGTGGAAAA TATACTCAAG ATGAAATTAA AAAGAAAATA AAAGGTAATG
      CD0113-buk  TAGTGGAAAA TATACTCAAG ATGAAATTAA AAAGAAAATA AAAGGTAATG

pE_CD0113-insert  GCGGACTAGT TGCATACTTA AACACTAATG ATGCTAGAGA AGTTGAAGAA
      CD0113-buk  GCGGATTAGT TGCATACTTA AACACTAATG ATGCTAGAGA AGTTGAAGAA

pE_CD0113-insert  AGAATTGAAG CTGGTGATGA AAAAGCTAAA TTAGTATATG AAGCTATGGC
      CD0113-buk  AGAATTGAAG CTGGTGATGA AAAAGCTAAA TTAGTATATG AAGCTATGGC

pE_CD0113-insert  ATATCAAATC TCTAAAGAAA TAGGAGCTAG TGCTGCAGTT CTTAAGGGAG
      CD0113-buk  ATATCAAATC TCTAAAGAAA TAGGAGCTAG TGCTGCAGTT CTTAAGGGAG

pE_CD0113-insert  ATGTAAAAGC AATATTATTA ACTGGTGGAA TCGCATATTC AAAAATGTTT
      CD0113-buk  ATGTAAAAGC AATATTATTA ACTGGTGGAA TCGCATATTC AAAAATGTTT
```

```
pE_CD0113-insert   ACAGAAATGA TTGCAGATAG AGTTAAATTT ATAGCAGATG TAAAAGTTTA
      CD0113-buk   ACAGAAATGA TTGCAGATAG AGTTAAATTT ATAGCAGATG TAAAAGTTTA

pE_CD0113-insert   TCCAGGTGAA GATGAAATGA TTGCATTAGC TCAAGGTGGA CTTAGAGTTT
      CD0113-buk   TCCAGGTGAA GATGAAATGA TTGCATTAGC TCAAGGTGGA CTTAGAGTTT

pE_CD0113-insert   TAACTGGTGA AGAAGAGGCT CAAGTTTATG ATAACGGGAG GAGACG
      CD0113-buk   TAACTGGTGA AGAAGAGGCT CAAGTTTATG ATAACTAA-- ------
```

App.32: Sequence alignment for the cloned *buk* from *C. difficile* DSMZ 1296^T and the genome derived genes from strain 630 [131] (GeneBank accession number CD0113). There are two mutations highlighted with red.

```
        pE_CD0113   MSKIFKILTI NPGSTSTKIA VFDNEDLVFE KTLRHSSEEI GKYEKVSDQF
               Buk   MSKIFKILTI NPGSTSTKIA VFDNEDLVFE KTLRHSSEEI GKYEKVSDQF

        pE_CD0113   EFRKQVIEEA LKEGGVKTSE LDAVVGRGGL LKPIKGGTYS VSAAMIEDLK
               Buk   EFRKQVIEEA LKEGGVKTSE LDAVVGRGGL LKPIKGGTYS VSAAMIEDLK

        pE_CD0113   VGVLGEHASN LGGIIAKQIG EEVNVPSYIV DPVVVDELED VARISGMPEI
               Buk   VGVLGEHASN LGGIIAKQIG EEVNVPSYIV DPVVVDELED VARISGMPEI

        pE_CD0113   SRASVVHALN QKAIARRYAR EINKKYEDIN LIVAHMGGGV SVGAHKNGKI
               Buk   SRASVVHALN QKAIARRYAR EINKKYEDIN LIVAHMGGGV SVGAHKNGKI

        pE_CD0113   VDVANALDGE GPFSPERSGG LPVGALVKMC FSGKYTQDEI KKKIKGNGGL
               Buk   VDVANALDGE GPFSPERSGG LPVGALVKMC FSGKYTQDEI KKKIKGNGGL

        pE_CD0113   VAYLNTNDAR EVEERIEAGD EKAKLVYEAM AYQISKEIGA SAAVLKGDVK
               Buk   VAYLNTNDAR EVEERIEAGD EKAKLVYEAM AYQISKEIGA SAAVLKGDVK

        pE_CD0113   AILLTGGIAY SKMFTEMIAD RVKFIADVKV YPGEDEMIAL AQGGLRVLTG
               Buk   AILLTGGIAY SKMFTEMIAD RVKFIADVKV YPGEDEMIAL AQGGLRVLTG

        pE_CD0113   EEEAQVYDNG
               Buk   EEEAQVYDN.
```

App.33: Amino acid alignment for the cloned *buk* from *C. difficile* DSMZ 1296^T and the genome derived genes from strain 630 [131]. No mutation was observed in the amino acid sequence. The stop codon was replaced with glycin.



Expected fragments (bp)	26+31	42	86	186	237	345	558	586	720

App.34: Analysis of *pE_buk (CD0113)* with MultiNA after digestion with *NlaIII* restriction enzyme. On the right hand, there is electropherogram for the plasmid. On the left hand, there is pDRAW file shows the recognition sites of *NlaIII* as well as the sequencing primers positions.

```
>pE_CA_C2711-insert
        1 CGTCTCCAAT GGATTTTAAT TTAACAAGAG AACAAGAATT AGTAAGACAG ATGGTTAGAG
       61 AATTTGCTGA AAATGAAGTT AAACCTATAG CAGCAGAAAT TGATGAAACA GAAAGATTTC
      121 CAATGGAAAA TGTAAAGAAA ATGGGTCAGT ATGGTATGAT GGGAATTCCA TTTTCAAAAG
      181 AGTATGGTGG CGCAGGTGGA GATGTATTAT CTTATATAAT CGCCGTTGAG GAATTATCAA
      241 AGGTTTGCGG TACTACAGGA GTTATTCTTT CAGCACATAC ATCACTTTGT GCTTCATTAA
      301 TAAATGAACA TGGTACAGAA GAACAAAAAC AAAAATATTT AGTACCTTTA GCTAAAGGTG
      361 AAAAAATAGG TGCTTATGGA TTGACTGAGC CAAATGCAGG AACAGATTCT GGAGCACAAC
      421 AAACAGTAGC TGTACTTGAA GGAGATCATT ATGTAATTAA TGGTTCAAAA ATATTCATAA
      481 CTAATGGAGG AGTTGCAGAT ACTTTTGTTA TATTTGCAAT GACTGACAGA ACTAAAGGAA
      541 CAAAAGGTAT ATCAGCATTT ATAATAGAAA AAGGCTTCAA AGGTTTCTCT ATTGGTAAAG
      601 TTGAACAAAA GCTTGGAATA AGAGCTTCAT CAACAACTGA ACTTGTATTT GAAGATATGA
      661 TAGTACCAGT AGAAAACATG ATTGGTAAAG AAGGAAAAGG CTTCCCTATA GCAATGAAAA
      721 CTCTTGATGG AGGAAGAATT GGTATAGCAG CTCAAGCTTT AGGTATAGCT GAAGGTGCTT
      781 TCAACGAAGC AAGAGCTTAC ATGAAGGAGA GAAAACAATT TGGAAGAAGC CTTGACAAAT
      841 TCCAAGGTCT TGCATGGATG ATGACAGATA TGGATGTAGC TATAGAATCA GCTAGATATT
      901 TAGTATATAA AGCAGCATAT CTTAAACAAG CAGGACTTCC ATACACAGTT GATGCTGCAA
      961 GAGCTAAGCT TCATGCTGCA AATGTAGCAA TGGATGTAAC AACTAAGGCA GTACAATTAT
     1021 TTGGTGGATA CGGATATACA AAAGATTATC CAGTTGAAAG AATGATGAGA GATGCTAAGA
     1081 TAACTGAAAT ATATGAAGGA ACTTCAGAAG TTCAGAAATT AGTTATTTCA GGAAAAATTT
     1141 TTAGAGGGAG GAGACG
```

App.35: The nucleotide sequence of *bcd* cloned from *C. acetobutylicum* DSMZ 792[T]. The gray highlighted sequences refer to the *Esp3I* restriction site. There are two mutations highlighted with red.

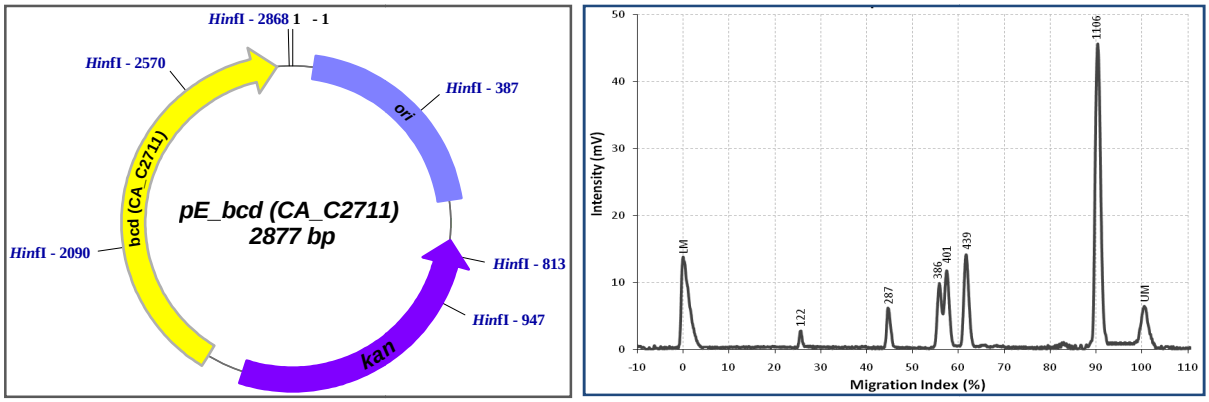
```
pE-CA_C2711-insert  CGTCTCCAAT GGATTTTAAT TTAACAAGAG AACAAGAATT AGTAAGACAG
      CA_C2711-bcd  --------AT GGATTTTAAT TTAACAAGAG AACAAGAATT AGTAAGACAG

pE-CA_C2711-insert  ATGGTTAGAG AATTTGCTGA AAATGAAGTT AAACCTATAG CAGCAGAAAT
      CA_C2711-bcd  ATGGTTAGAG AATTTGCTGA AAATGAAGTT AAACCTATAG CAGCAGAAAT

pE-CA_C2711-insert  TGATGAAACA GAAAGATTTC CAATGGAAAA TGTAAAGAAA ATGGGTCAGT
      CA_C2711-bcd  TGATGAAACA GAAAGATTTC CAATGGAAAA TGTAAAGAAA ATGGGTCAGT

pE-CA_C2711-insert  ATGGTATGAT GGGAATTCCA TTTTCAAAAG AGTATGGTGG CGCAGGTGGA
      CA_C2711-bcd  ATGGTATGAT GGGAATTCCA TTTTCAAAAG AGTATGGTGG CGCAGGTGGA

pE-CA_C2711-insert  GATGTATTAT CTTATATAAT CGCCGTTGAG GAATTATCAA AGGTTTGCGG
      CA_C2711-bcd  GATGTATTAT CTTATATAAT CGCCGTTGAG GAATTATCAA AGGTTTGCGG

pE-CA_C2711-insert  TACTACAGGA GTTATTCTTT CAGCACATAC ATCACTTTGT GCTTCATTAA
      CA_C2711-bcd  TACTACAGGA GTTATTCTTT CAGCACATAC ATCACTTTGT GCTTCATTAA

pE-CA_C2711-insert  TAAATGAACA TGGTACAGAA GAACAAAAAC AAAAATATTT AGTACCTTTA
      CA_C2711-bcd  TAAATGAACA TGGTACAGAA GAACAAAAAC AAAAATATTT AGTACCTTTA

pE-CA_C2711-insert  GCTAAAGGTG AAAAAATAGG TGCTTATGGA TTGACTGAGC CAAATGCAGG
      CA_C2711-bcd  GCTAAAGGTG AAAAAATAGG TGCTTATGGA TTGACTGAGC CAAATGCAGG

pE-CA_C2711-insert  AACAGATTCT GGAGCACAAC AAACAGTAGC TGTACTTGAA GGAGATCATT
      CA_C2711-bcd  AACAGATTCT GGAGCACAAC AAACAGTAGC TGTACTTGAA GGAGATCATT

pE-CA_C2711-insert  ATGTAATTAA TGGTTCAAAA ATATTCATAA CTAATGGAGG AGTTGCAGAT
      CA_C2711-bcd  ATGTAATTAA TGGTTCAAAA ATATTCATAA CTAATGGAGG AGTTGCAGAT

pE-CA_C2711-insert  ACTTTTGTTA TATTTGCAAT GACTGACAGA ACTAAAGGAA CAAAAGGTAT
      CA_C2711-bcd  ACTTTTGTTA TATTTGCAAT GACTGACAGA ACTAAAGGAA CAAAAGGTAT

pE-CA_C2711-insert  ATCAGCATTT ATAATAGAAA AAGGCTTCAA AGGTTTCTCT ATTGGTAAAG
      CA_C2711-bcd  ATCAGCATTT ATAATAGAAA AAGGCTTCAA AGGTTTCTCT ATTGGTAAAG

pE-CA_C2711-insert  TTGAACAAAA GCTTGGAATA AGAGCTTCAT CAACAACTGA ACTTGTATTT
      CA_C2711-bcd  TTGAACAAAA GCTTGGAATA AGAGCTTCAT CAACAACTGA ACTTGTATTT
```

```
pE-CA_C2711-insert   GAAGATATGA TAGTACCAGT AGAAAACATG ATTGGTAAAG AAGGAAAAGG
      CA_C2711-bcd   GAAGATATGA TAGTACCAGT AGAAAACATG ATTGGTAAAG AAGGAAAAGG

pE-CA_C2711-insert   CTTCCCTATA GCAATGAAAA CTCTTGATGG AGGAAGAATT GGTATAGCAG
      CA_C2711-bcd   CTTCCCTATA GCAATGAAAA CTCTTGATGG AGGAAGAATT GGTATAGCAG

pE-CA_C2711-insert   CTCAAGCTTT AGGTATAGCT GAAGGTGCTT TCAACGAAGC AAGAGCTTAC
      CA_C2711-bcd   CTCAAGCTTT AGGTATAGCT GAAGGTGCTT TCAACGAAGC AAGAGCTTAC

pE-CA_C2711-insert   ATGAAGGAGA GAAAACAATT TGGAAGAAGC CTTGACAAAT TCCAAGGTCT
      CA_C2711-bcd   ATGAAGGAGA GAAAACAATT TGGAAGAAGC CTTGACAAAT TCCAAGGTCT

pE-CA_C2711-insert   TGCATGGATG ATGACAGATA TGGATGTAGC TATAGAATCA GCTAGATATT
      CA_C2711-bcd   TGCATGGATG ATGGCAGATA TGGATGTAGC TATAGAATCA GCTAGATATT

pE-CA_C2711-insert   TAGTATATAA AGCAGCATAT CTTAAACAAG CAGGACTTCC ATACACAGTT
      CA_C2711-bcd   TAGTATATAA AGCAGCATAT CTTAAACAAG CAGGACTTCC ATACACAGTT

pE-CA_C2711-insert   GATGCTGCAA GAGCTAAGCT TCATGCTGCA AATGTAGCAA TGGATGTAAC
      CA_C2711-bcd   GATGCTGCAA GAGCTAAGCT TCATGCTGCA AATGTAGCAA TGGATGTAAC

pE-CA_C2711-insert   AACTAAGGCA GTACAATTAT TTGGTGGATA CGGATATACA AAAGATTATC
      CA_C2711-bcd   AACTAAGGCA GTACAATTAT TTGGTGGATA CGGATATACA AAAGATTATC

pE-CA_C2711-insert   CAGTTGAAAG AATGATGAGA GATGCTAAGA TAACTGAAAT ATATGAAGGA
      CA_C2711-bcd   CAGTTGAAAG AATGATGAGA GATGCTAAGA TAACTGAAAT ATATGAAGGA

pE-CA_C2711-insert   ACTTCAGAAG TTCAGAAATT AGTTATTTCA GGAAAAATTT TTAGAGGGAG
      CA_C2711-bcd   ACTTCAGAAG TTCAGAAATT AGTTATTTCA GGAAAAATTT TTAGATAA-

pE-CA_C2711-insert   GAGACG
      CA_C2711-bcd   ------
```

App.36: Sequence alignment for the cloned *bcd* from *C. acetobutylicum* DSMZ 792^T and the genome derived genes from strain ATCC 824 (CA_C2711) [107]. One mutation was observed.

```
    pE-CA_C2711   MDFNLTREQE LVRQMVREFA ENEVKPIAAE IDETERFPME NVKKMGQYGM
            Bcd   MDFNLTREQE LVRQMVREFA ENEVKPIAAE IDETERFPME NVKKMGQYGM

    pE-CA_C2711   MGIPFSKEYG GAGGDVLSYI IAVEELSKVC GTTGVILSAH TSLCASLINE
            Bcd   MGIPFSKEYG GAGGDVLSYI IAVEELSKVC GTTGVILSAH TSLCASLINE

    pE-CA_C2711   HGTEEQKQKY LVPLAKGEKI GAYGLTEPNA GTDSGAQQTV AVLEGDHYVI
            Bcd   HGTEEQKQKY LVPLAKGEKI GAYGLTEPNA GTDSGAQQTV AVLEGDHYVI

    pE-CA_C2711   NGSKIFITNG GVADTFVIFA MTDRTKGTKG ISAFIIEKGF KGFSIGKVEQ
            Bcd   NGSKIFITNG GVADTFVIFA MTDRTKGTKG ISAFIIEKGF KGFSIGKVEQ

    pE-CA_C2711   KLGIRASSTT ELVFEDMIVP VENMIGKEGK GFPIAMKTLD GGRIGIAAQA
            Bcd   KLGIRASSTT ELVFEDMIVP VENMIGKEGK GFPIAMKTLD GGRIGIAAQA

    pE-CA_C2711   LGIAEGAFNE ARAYMKERKQ FGRSLDKFQG LAWMMTDMDV AIESARYLVY
            Bcd   LGIAEGAFNE ARAYMKERKQ FGRSLDKFQG LAWMMADMDV AIESARYLVY

    pE-CA_C2711   KAAYLKQAGL PYTVDAARAK LHAANVAMDV TTKAVQLFGG YGYTKDYPVE
            Bcd   KAAYLKQAGL PYTVDAARAK LHAANVAMDV TTKAVQLFGG YGYTKDYPVE

    pE-CA_C2711   RMMRDAKITE IYEGTSEVQK LVISGKIFRG
            Bcd   RMMRDAKITE IYEGTSEVQK LVISGKIFR.
```

App.37: Amino acid alignment for the cloned *bcd* from *C. acetobutylicum* DSMZ 792^T and the genome derived genes from strain ATCC 824 (CA_C2711) [107]. Only one conservative amino acid exchange was observed, which is also found in the donor strain. The stop codon was replaced with glycin.



Expected fragments size (bp)	134	298	396	426	480	1143

App.38: Analysis of *pE_bcd* (CA_C2711) with MultiNA after digestion with *HinfI* restriction enzyme. On the right hand, there is electropherogram for the plasmid. On the left hand, there is pDRAW file shows the recognition sites of *HinfI* as well as the sequencing primers positions.

>pE_CA_C2710-insert

```
       1 CGTCTCCAAT GAATATAGTT GTTTGTTTAA AACAAGTTCC AGATACAGCG GAAGTTAGAA
      61 TAGATCCAGT TAAGGGAACA CTTATAAGAG AAGGAGTTCC ATCAATAATA AATCCAGATG
     121 ATAAAAACGC ACTTGAGGAA GCTTTAGTAT TAAAAGATAA TTATGGTGCA CATGTAACAG
     181 TTATAAGTAT GGGACCTCCA CAAGCTAAAA ATGCTTTAGT AGAAGCTTTG GCTATGGGTG
     241 CTGATGAAGC TGTACTTTTA ACAGATAGAG CATTTGGAGG AGCAGATACA CTTGCGACTT
     301 CACATACAAT TGCAGCAGGA ATTAAGAAGC TAAAATATGA TATAGTTTTT GCTGGAAGGC
     361 AGGCTATAGA TGGAGATACA GCTCAGGTTG GACCAGAAAT AGCTGAGCAT CTTGGAATAC
     421 CTCAAGTAAC TTATGTTGAG AAAGTTGAAG TTGATGGAGA TACTTTAAAG ATTAGAAAAG
     481 CTTGGGAAGA TGGATATGAA GTTGTTGAAG TTAAGACACC AGTTCTTTTA ACAGCAATTA
     541 AAGAATTAAA TGTTCCAAGA TATATGAGTG TAGAAAAAAT ATTCGGAGCA TTTGATAAAG
     601 AAGTAAAAAT GTGGACTGCC GATGATATAG ATGTAGATAA GGCTAATTTA GGTCTTAAAG
     661 GTTCACCAAC TAAAGTTAAG AAGTCATCAA CTAAAGAAGT TAAAGGACAG GGAGAAGTTA
     721 TTGATAAGCC TGTTAAGGAA GCAGCTGCAT ATGTTGTCTC AAAATTAAAA GAAGAACACT
     781 ATATTGGGAG GAGACG
```

App.39: The nucleotide sequence of *etfB* cloned from *C. acetobutylicum* DSMZ 792^{T}. The gray highlighted sequences refer to the *Esp3I* restriction site. No difference was obtained.

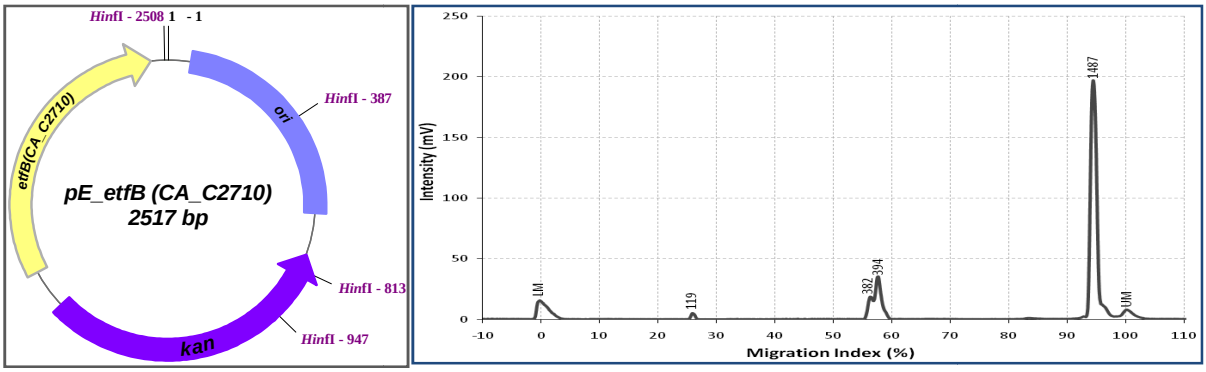
```
pE-CA_C2710-insert  CGTCTCCAAT GAATATAGTT GTTTGTTTAA AACAAGTTCC AGATACAGCG
     CA_C2710-etfB  --------AT GAATATAGTT GTTTGTTTAA AACAAGTTCC AGATACAGCG

pE-CA_C2710-insert  GAAGTTAGAA TAGATCCAGT TAAGGGAACA CTTATAAGAG AAGGAGTTCC
     CA_C2710-etfB  GAAGTTAGAA TAGATCCAGT TAAGGGAACA CTTATAAGAG AAGGAGTTCC

pE-CA_C2710-insert  ATCAATAATA AATCCAGATG ATAAAAACGC ACTTGAGGAA GCTTTAGTAT
     CA_C2710-etfB  ATCAATAATA AATCCAGATG ATAAAAACGC ACTTGAGGAA GCTTTAGTAT

pE-CA_C2710-insert  TAAAAGATAA TTATGGTGCA CATGTAACAG TTATAAGTAT GGGACCTCCA
     CA_C2710-etfB  TAAAAGATAA TTATGGTGCA CATGTAACAG TTATAAGTAT GGGACCTCCA

pE-CA_C2710-insert  CAAGCTAAAA ATGCTTTAGT AGAAGCTTTG GCTATGGGTG CTGATGAAGC
     CA_C2710-etfB  CAAGCTAAAA ATGCTTTAGT AGAAGCTTTG GCTATGGGTG CTGATGAAGC

pE-CA_C2710-insert  TGTACTTTTA ACAGATAGAG CATTTGGAGG AGCAGATACA CTTGCGACTT
     CA_C2710-etfB  TGTACTTTTA ACAGATAGAG CATTTGGAGG AGCAGATACA CTTGCGACTT

pE-CA_C2710-insert  CACATACAAT TGCAGCAGGA ATTAAGAAGC TAAAATATGA TATAGTTTTT
     CA_C2710-etfB  CACATACAAT TGCAGCAGGA ATTAAGAAGC TAAAATATGA TATAGTTTTT
```

```
pE-CA_C2710-insert   GCTGGAAGGC AGGCTATAGA TGGAGATACA GCTCAGGTTG GACCAGAAAT
     CA_C2710-etfB   GCTGGAAGGC AGGCTATAGA TGGAGATACA GCTCAGGTTG GACCAGAAAT

pE-CA_C2710-insert   AGCTGAGCAT CTTGGAATAC CTCAAGTAAC TTATGTTGAG AAAGTTGAAG
     CA_C2710-etfB   AGCTGAGCAT CTTGGAATAC CTCAAGTAAC TTATGTTGAG AAAGTTGAAG

pE-CA_C2710-insert   TTGATGGAGA TACTTTAAAG ATTAGAAAAG CTTGGGAAGA TGGATATGAA
     CA_C2710-etfB   TTGATGGAGA TACTTTAAAG ATTAGAAAAG CTTGGGAAGA TGGATATGAA

pE-CA_C2710-insert   GTTGTTGAAG TTAAGACACC AGTTCTTTTA ACAGCAATTA AAGAATTAAA
     CA_C2710-etfB   GTTGTTGAAG TTAAGACACC AGTTCTTTTA ACAGCAATTA AAGAATTAAA

pE-CA_C2710-insert   TGTTCCAAGA TATATGAGTG TAGAAAAAAT ATTCGGAGCA TTTGATAAAG
     CA_C2710-etfB   TGTTCCAAGA TATATGAGTG TAGAAAAAAT ATTCGGAGCA TTTGATAAAG

pE-CA_C2710-insert   AAGTAAAAAT GTGGACTGCC GATGATATAG ATGTAGATAA GGCTAATTTA
     CA_C2710-etfB   AAGTAAAAAT GTGGACTGCC GATGATATAG ATGTAGATAA GGCTAATTTA

pE-CA_C2710-insert   GGTCTTAAAG GTTCACCAAC TAAAGTTAAG AAGTCATCAA CTAAAGAAGT
     CA_C2710-etfB   GGTCTTAAAG GTTCACCAAC TAAAGTTAAG AAGTCATCAA CTAAAGAAGT

pE-CA_C2710-insert   TAAAGGACAG GGAGAAGTTA TTGATAAGCC TGTTAAGGAA GCAGCTGCAT
     CA_C2710-etfB   TAAAGGACAG GGAGAAGTTA TTGATAAGCC TGTTAAGGAA GCAGCTGCAT

pE-CA_C2710-insert   ATGTTGTCTC AAAATTAAAA GAAGAACACT ATATTGGGAG GAGACG
     CA_C2710-etfB   ATGTTGTCTC AAAATTAAAA GAAGAACACT ATATTTAA-- ------
```

App.40: Sequence alignment for the cloned *etfB* from *C. acetobutylicum* DSMZ 792^{T} and the genome derived genes from strain ATCC 824 (CA_C2710) [107]. No mutation was observed.

```
    pE-CA_C2710  MNIVVCLKQV PDTAEVRIDP VKGTLIREGV PSIINPDDKN ALEEALVLKD
           etfB  MNIVVCLKQV PDTAEVRIDP VKGTLIREGV PSIINPDDKN ALEEALVLKD

    pE-CA_C2710  NYGAHVTVIS MGPPQAKNAL VEALAMGADE AVLLTDRAFG GADTLATSHT
           etfB  NYGAHVTVIS MGPPQAKNAL VEALAMGADE AVLLTDRAFG GADTLATSHT

    pE-CA_C2710  IAAGIKKLKY DIVFAGRQAI DGDTAQVGPE IAEHLGIPQV TYVEKVEVDG
           etfB  IAAGIKKLKY DIVFAGRQAI DGDTAQVGPE IAEHLGIPQV TYVEKVEVDG

    pE-CA_C2710  DTLKIRKAWE DGYEVVEVKT PVLLTAIKEL NVPRYMSVEK IFGAFDKEVK
           etfB  DTLKIRKAWE DGYEVVEVKT PVLLTAIKEL NVPRYMSVEK IFGAFDKEVK

    pE-CA_C2710  MWTADDIDVD KANLGLKGSP TKVKKSSTKE VKGQGEVIDK PVKEAAAYVV
           etfB  MWTADDIDVD KANLGLKGSP TKVKKSSTKE VKGQGEVIDK PVKEAAAYVV

    pE-CA_C2710  SKLKEEHYIG
           etfB  SKLKEEHYI.
```

App.41: Amino acid alignment for the cloned *etfB* from *C. acetobutylicum* DSMZ 792^{T} and the genome derived genes from strain ATCC 824 (CA_C2710) [107]. No mutation was observed. The stop codon was replaced with glycin.



Expected fragments sizes (bp)	134	396	426	1561

App.42: Analysis of digest *pE_etfB(CA_C2710)* with MultiNA after digestion with *HinfI* restriction enzyme. On the right hand, there is electropherogram for the plasmid. On the left hand, there is pDRAW file shows the recognition sites of *HinfI* as well as the sequencing primers positions.

```
>pE_CA_C2709-insert
        1 CGTCTCCAAT GAATAAAGCA GATTACAAGG GCGTATGGGT GTTTGCTGAA CAAAGAGATG
       61 GAGAATTACA AAAGGTATCA TTGGAATTAT TAGGTAAAGG TAAGGAAATG GCTGAGAAAT
      121 TAGGCGTTGA ATTAACAGCT GTTTTACTTG GACATAATAC TGAAAAAATG TCAAAGGATT
      181 TATTATCTCA TGGAGCAGAT AAGGTTTTAG CAGCAGATAA TGAACTTTTA GCACATTTTT
      241 CAACAGATGG ATATGCTAAA GTTATATGTG ATTTAGTTAA TGAAAGAAAG CCAGAAATAT
      301 TATTCATAGG AGCTACTTTC ATAGGAAGAG ATTTAGGACC AAGAATAGCA GCAAGACTTT
      361 CTACTGGTTT AACTGCTGAT TGTACATCAC TTGACATAGA TGTAGAAAAT AGAGATTTAT
      421 TGGCTACAAG ACCAGCGTTT GGTGGAAATT TGATAGCTAC AATAGTTTGT TCAGACCACA
      481 GACCACAAAT GGCTACAGTA AGACCTGGTG TGTTTGAAAA ATTACCTGTT AATGATGCAA
      541 ATGTTTCTGA TGATAAAATA GAAAAAGTTG CAATTAAATT AACAGCATCA GACATAAGAA
      601 CAAAAGTTTC AAAAGTTGTT AAGCTTGCTA AAGATATTGC AGATATCGGA GAAGCTAAGG
      661 TATTAGTTGC TGGTGGTAGA GGAGTTGGAA GCAAAGAAAA CTTTGAAAAA CTTGAAGAGT
      721 TAGCAAGTTT ACTTGGTGGA ACAATAGCCG CTTCAAGAGC AGCAATAGAA AAAGAATGGG
      781 TTGATAAGGA CCTTCAAGTA GGTCAAACTG GTAAAACTGT AAGACCAACT CTTTATATTG
      841 CATGTGGTAT ATCAGGAGCT ATCCAGCATT TAGCAGGTAT GCAAGATTCA GATTACATAA
      901 TTGCTATAAA TAAAGATGTA GAAGCCCCAA TAATGAAGGT AGCAGATTTG GCTATAGTTG
      961 GTGATGTAAA TAAAGTTGTA CCAGAATTAA TAGCTCAAGT TAAAGCTGCT AATAATGGGA
     1021 GGAGACG
```

App.43: The nucleotide sequence of *etfA* cloned from *C. acetobutylicum* DSMZ 792^{T}. The gray highlighted sequences refer to the *Esp3I* restriction site. A silent mutation was introduced to remove the *LguI* restriction site and it is green highlighted.

```
pE-CA_C2709-insert  CGTCTCCAAT GAATAAAGCA GATTACAAGG GCGTATGGGT GTTTGCTGAA
     CA_C2709-etfA  --------AT GAATAAAGCA GATTACAAGG GCGTATGGGT GTTTGCTGAA

pE-CA_C2709-insert  CAAAGAGATG GAGAATTACA AAAGGTATCA TTGGAATTAT TAGGTAAAGG
     CA_C2709-etfA  CAAAGAGACG GAGAATTACA AAAGGTATCA TTGGAATTAT TAGGTAAAGG

pE-CA_C2709-insert  TAAGGAAATG GCTGAGAAAT TAGGCGTTGA ATTAACAGCT GTTTTACTTG
     CA_C2709-etfA  TAAGGAAATG GCTGAGAAAT TAGGCGTTGA ATTAACAGCT GTTTTACTTG

pE-CA_C2709-insert  GACATAATAC TGAAAAAATG TCAAAGGATT TATTATCTCA TGGAGCAGAT
     CA_C2709-etfA  GACATAATAC TGAAAAAATG TCAAAGGATT TATTATCTCA TGGAGCAGAT

pE-CA_C2709-insert  AAGGTTTTAG CAGCAGATAA TGAACTTTTA GCACATTTTT CAACAGATGG
     CA_C2709-etfA  AAGGTTTTAG CAGCAGATAA TGAACTTTTA GCACATTTTT CAACAGATGG

pE-CA_C2709-insert  ATATGCTAAA GTTATATGTG ATTTAGTTAA TGAAAGAAAG CCAGAAATAT
     CA_C2709-etfA  ATATGCTAAA GTTATATGTG ATTTAGTTAA TGAAAGAAAG CCAGAAATAT
```

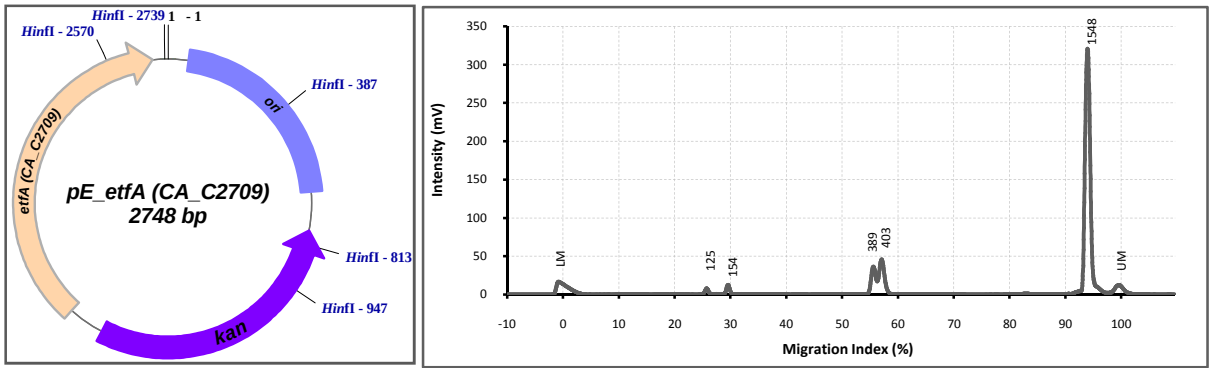
```
pE-CA_C2709-insert   TATTCATAGG AGCTACTTTC ATAGGAAGAG ATTTAGGACC AAGAATAGCA
     CA_C2709-etfA   TATTCATAGG AGCTACTTTC ATAGGAAGAG ATTTAGGACC AAGAATAGCA

pE-CA_C2709-insert   GCAAGACTTT CTACTGGTTT AACTGCTGAT TGTACATCAC TTGACATAGA
     CA_C2709-etfA   GCAAGACTTT CTACTGGTTT AACTGCTGAT TGTACATCAC TTGACATAGA

pE-CA_C2709-insert   TGTAGAAAAT AGAGATTTAT TGGCTACAAG ACCAGCGTTT GGTGGAAATT
     CA_C2709-etfA   TGTAGAAAAT AGAGATTTAT TGGCTACAAG ACCAGCGTTT GGTGGAAATT

pE-CA_C2709-insert   TGATAGCTAC AATAGTTTGT TCAGACCACA GACCACAAAT GGCTACAGTA
     CA_C2709-etfA   TGATAGCTAC AATAGTTTGT TCAGACCACA GACCACAAAT GGCTACAGTA

pE-CA_C2709-insert   AGACCTGGTG TGTTTGAAAA ATTACCTGTT AATGATGCAA ATGTTTCTGA
     CA_C2709-etfA   AGACCTGGTG TGTTTGAAAA ATTACCTGTT AATGATGCAA ATGTTTCTGA

pE-CA_C2709-insert   TGATAAAATA GAAAAAGTTG CAATTAAATT AACAGCATCA GACATAAGAA
     CA_C2709-etfA   TGATAAAATA GAAAAAGTTG CAATTAAATT AACAGCATCA GACATAAGAA

pE-CA_C2709-insert   CAAAAGTTTC AAAAGTTGTT AAGCTTGCTA AAGATATTGC AGATATCGGA
     CA_C2709-etfA   CAAAAGTTTC AAAAGTTGTT AAGCTTGCTA AAGATATTGC AGATATCGGA

pE-CA_C2709-insert   GAAGCTAAGG TATTAGTTGC TGGTGGTAGA GGAGTTGGAA GCAAAGAAAA
     CA_C2709-etfA   GAAGCTAAGG TATTAGTTGC TGGTGGTAGA GGAGTTGGAA GCAAAGAAAA

pE-CA_C2709-insert   CTTTGAAAAA CTTGAAGAGT TAGCAAGTTT ACTTGGTGGA ACAATAGCCG
     CA_C2709-etfA   CTTTGAAAAA CTTGAAGAGT TAGCAAGTTT ACTTGGTGGA ACAATAGCCG

pE-CA_C2709-insert   CTTCAAGAGC AGCAATAGAA AAAGAATGGG TTGATAAGGA CCTTCAAGTA
     CA_C2709-etfA   CTTCAAGAGC AGCAATAGAA AAAGAATGGG TTGATAAGGA CCTTCAAGTA

pE-CA_C2709-insert   GGTCAAACTG GTAAAACTGT AAGACCAACT CTTTATATTG CATGTGGTAT
     CA_C2709-etfA   GGTCAAACTG GTAAAACTGT AAGACCAACT CTTTATATTG CATGTGGTAT

pE-CA_C2709-insert   ATCAGGAGCT ATCCAGCATT TAGCAGGTAT GCAAGATTCA GATTACATAA
     CA_C2709-etfA   ATCAGGAGCT ATCCAGCATT TAGCAGGTAT GCAAGATTCA GATTACATAA

pE-CA_C2709-insert   TTGCTATAAA TAAAGATGTA GAAGCCCCAA TAATGAAGGT AGCAGATTTG
     CA_C2709-etfA   TTGCTATAAA TAAAGATGTA GAAGCCCCAA TAATGAAGGT AGCAGATTTG

pE-CA_C2709-insert   GCTATAGTTG GTGATGTAAA TAAAGTTGTA CCAGAATTAA TAGCTCAAGT
     CA_C2709-etfA   GCTATAGTTG GTGATGTAAA TAAAGTTGTA CCAGAATTAA TAGCTCAAGT

pE-CA_C2709-insert   TAAAGCTGCT AATAATGGGA GGAGACG
     CA_C2709-etfA   TAAAGCTGCT AATAATTAA- -------
```

App.44: Sequence alignment for the cloned *etfA* from *C. acetobutylicum* DSMZ 792^{T} and the genome derived genes from strain ATCC 824 (CA_C2709) [107]. A silent mutation was introduced to remove the *LguI* restriction site and it is green highlighted

```
    pE-CA_C2709   MNKADYKGVW VFAEQRDGEL QKVSLELLGK GKEMAEKLGV ELTAVLLGHN
          etfA    MNKADYKGVW VFAEQRDGEL QKVSLELLGK GKEMAEKLGV ELTAVLLGHN

    pE-CA_C2709   TEKMSKDLLS HGADKVLAAD NELLAHFSTD GYAKVICDLV NERKPEILFI
          etfA    TEKMSKDLLS HGADKVLAAD NELLAHFSTD GYAKVICDLV NERKPEILFI

    pE-CA_C2709   GATFIGRDLG PRIAARLSTG LTADCTSLDI DVENRDLLAT RPAFGGNLIA
          etfA    GATFIGRDLG PRIAARLSTG LTADCTSLDI DVENRDLLAT RPAFGGNLIA

    pE-CA_C2709   TIVCSDHRPQ MATVRPGVFE KLPVNDANVS DDKIEKVAIK LTASDIRTKV
          etfA    TIVCSDHRPQ MATVRPGVFE KLPVNDANVS DDKIEKVAIK LTASDIRTKV
```

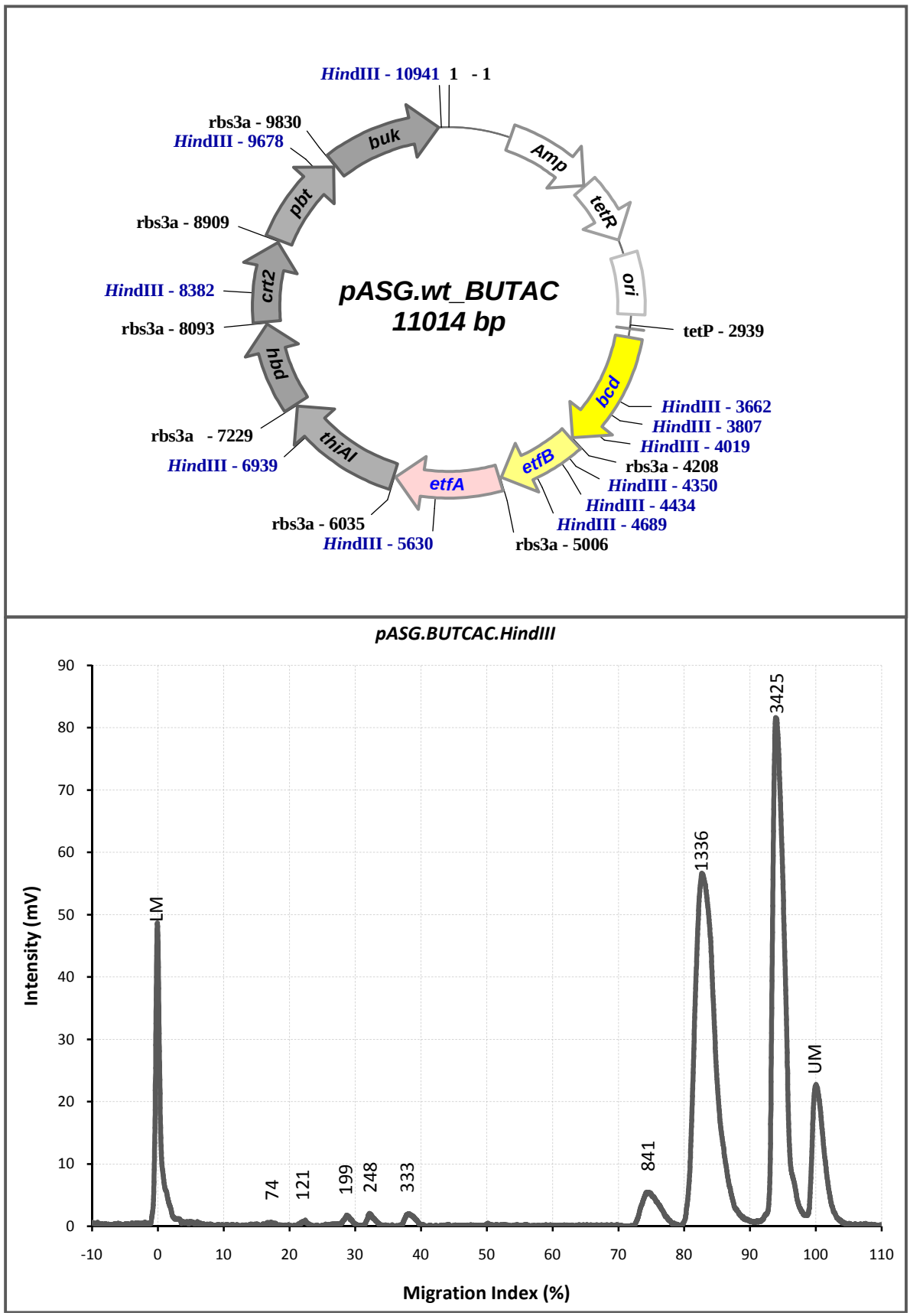
```
pE-CA_C2709  SKVVKLAKDI ADIGEAKVLV AGGRGVGSKE NFEKLEELAS LLGGTIAASR
       etfA  SKVVKLAKDI ADIGEAKVLV AGGRGVGSKE NFEKLEELAS LLGGTIAASR

pE-CA_C2709  AAIEKEWVDK DLQVGQTGKT VRPTLYIACG ISGAIQHLAG MQDSDYIIAI
       etfA  AAIEKEWVDK DLQVGQTGKT VRPTLYIACG ISGAIQHLAG MQDSDYIIAI

pE-CA_C2709  NKDVEAPIMK VADLAIVGDV NKVVPELIAQ VKAANNG
       etfA  NKDVEAPIMK VADLAIVGDV NKVVPELIAQ VKAANN.
```

App.45: Amino acid alignment for the cloned *etfA* from *C. acetobutylicum* DSMZ 792^{T} and the genome derived genes from strain ATCC 824 (CA_C2709) [107]. No mutation was observed. The stop codon was replaced with glycin.



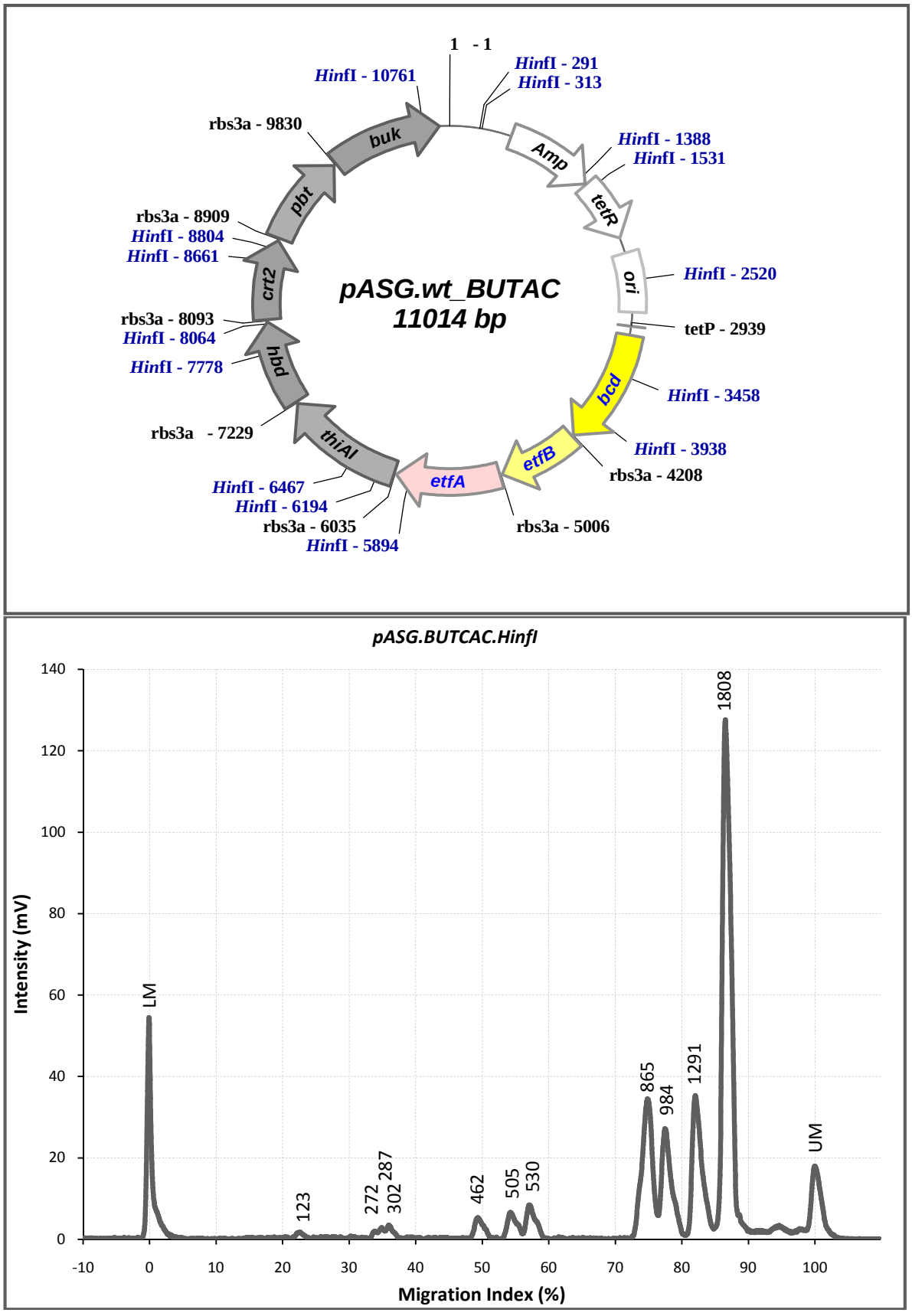
Expected fragments size (bp)	134	169	396	426	1623

App.46: Analysis of digest *pE_etfA(CA_C2709)* with MultiNA after digestion with *HinfI*.



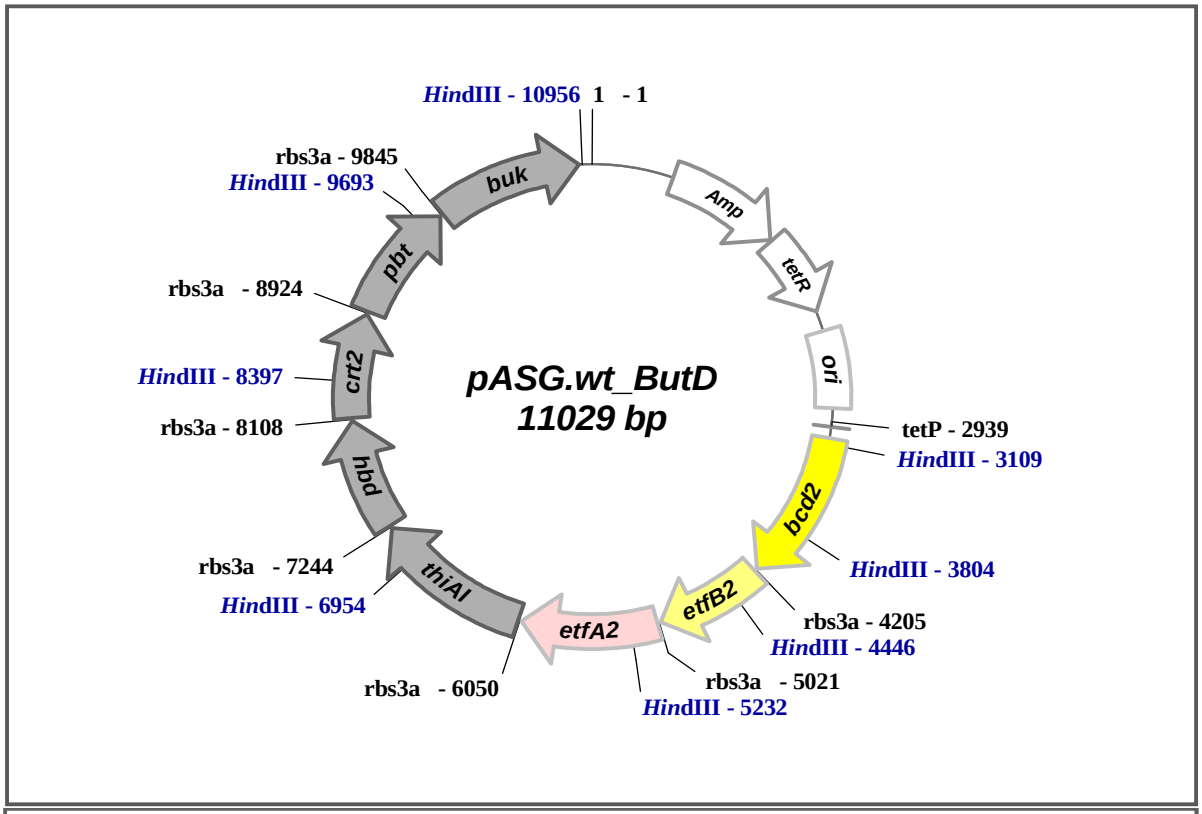
Expected fragments (bp)	84	145	212	255	331	941	1263+1296+1309+1443	3735

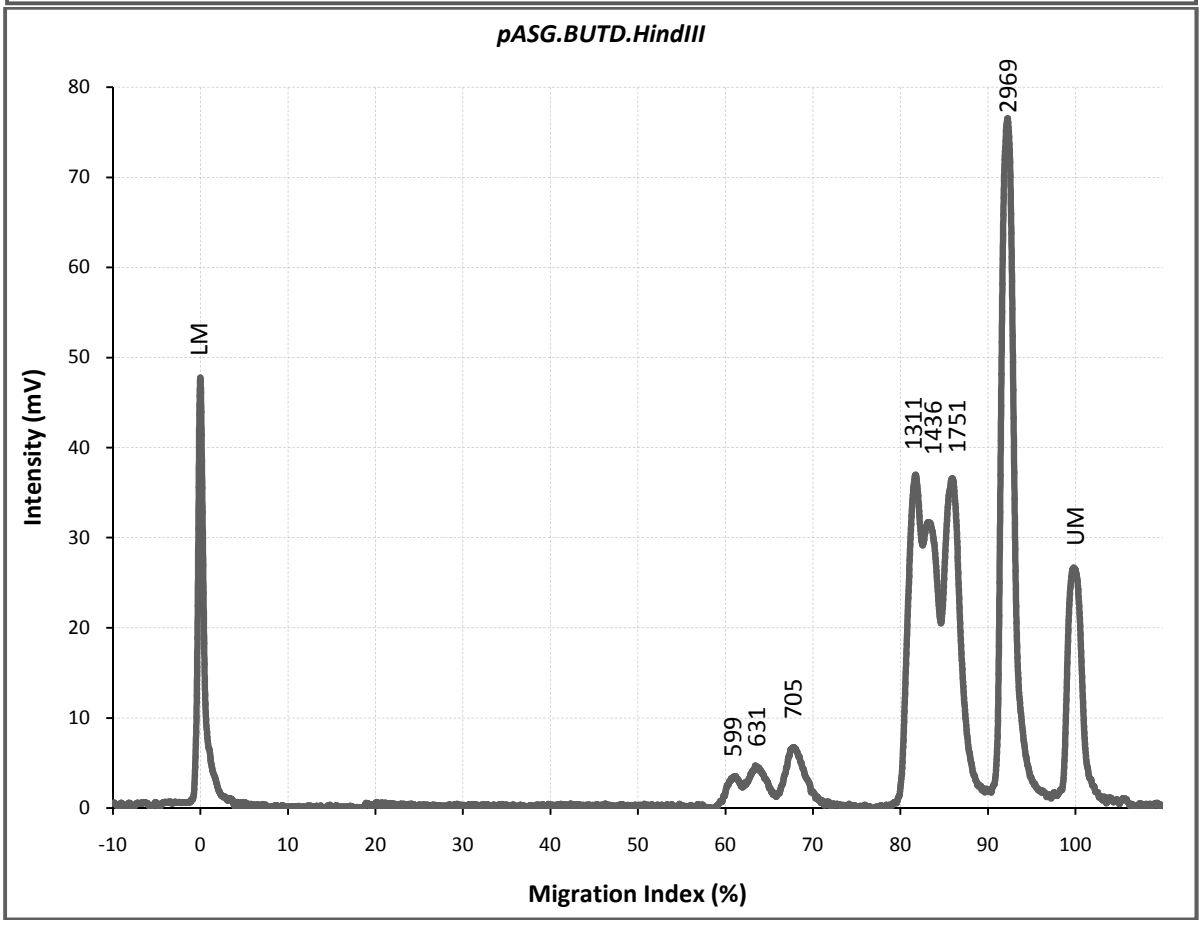
App.47: Analysis of digest *pASG_BUT$_{AC}$* with MultiNA after digestion with *HindIII* restriction enzyme. The above diagram is a pDRAW file shows the recognition sites of *HindIII* and the below one is electropherogram for the digested plasmid.



Expected fragments (bp)	143 +143	273	286	300	480	544	597	938+ 989	1075	1311	1956+ 1957

App.48: Analyze of digest *pASG_BUT$_{AC}$* with MultiNA after digestion with *HinfI* restriction enzyme. The above diagram is a pDRAW file shows the recognition sites of *HinfI* and the below one is electropherogram for the digested plasmid.

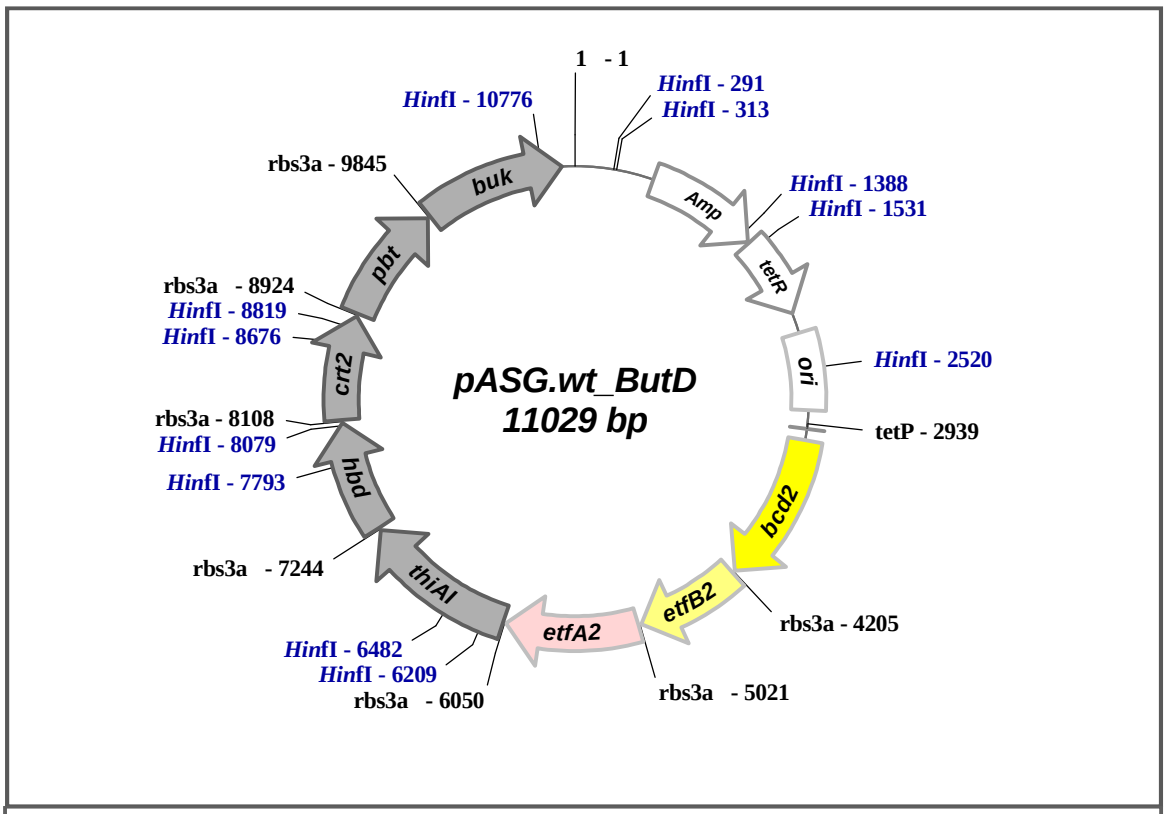


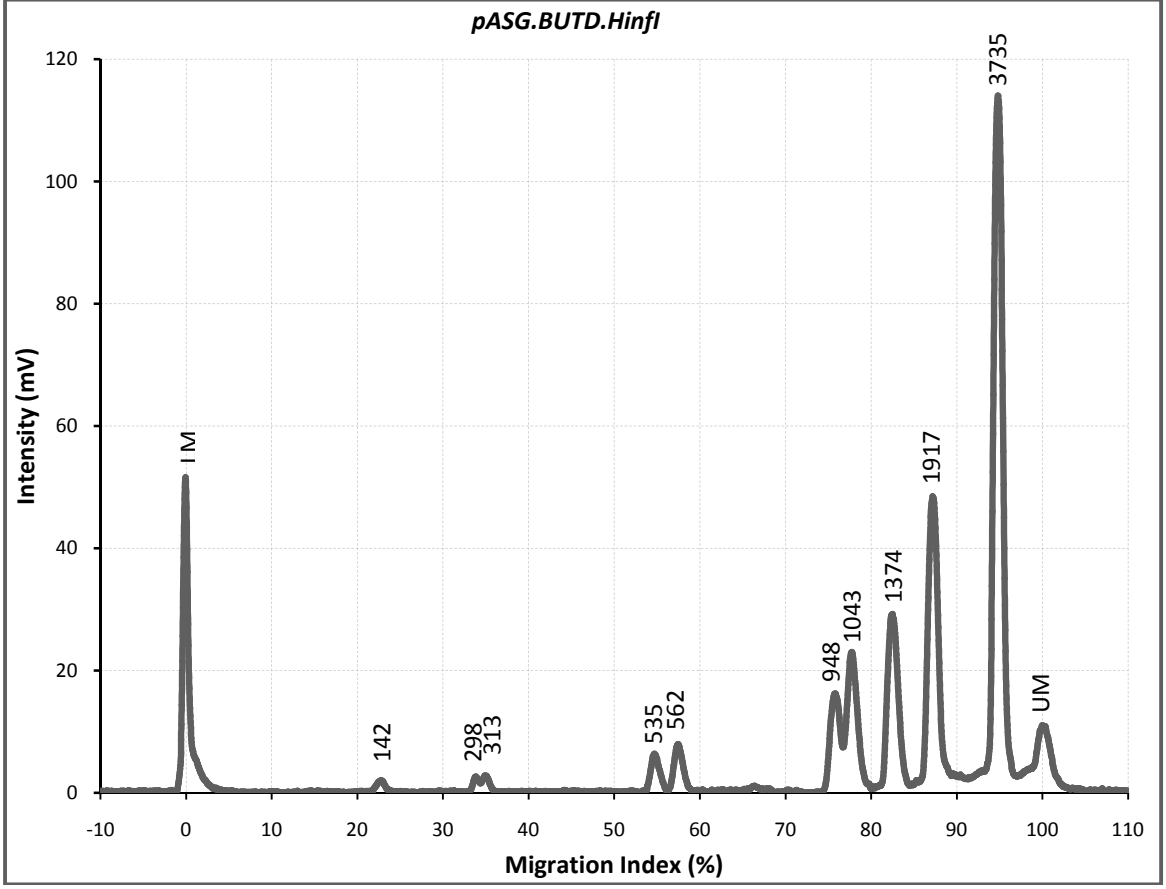


Expected fragments (bp)	642	659	786	1263+1296	1443	1722	3182

App.49: Analysis of digest *pASG_BUT$_D$* with MultiNA after digestion with *HindIII* restriction enzyme.

The above diagram is a pDRAW file shows the recognition sites of *HindIII* and the below one is electropherogram for the digested plasmid.





Expected fragments (bp)	143 +143	273	286	544	597	989	1075	1311	1956	3689

App.50: Analysis of digest *pASG_BUT$_D$* with MultiNA after digestion with *HinfI* restriction enzyme. The above diagram is a pDRAW file shows the recognition sites of *HinfI* and the below one is electropherogram for the digested plasmid.